ANFIS BASED DATA RATE PREDICTION FOR COGNITIVE RADIO

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIRMENTS FOR THE DEGREE OF

MASTER OF TECHNOLOGY
IN

TELEMATICS AND SIGNAL PROCESSING

BY

SHRISHAILAYYA M HIREMATH
ROLL NO: 208EC110

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA, INDIA

2010

ANFIS BASED DATA RATE PREDICTION FOR COGNITIVE RADIO

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIRMENTS FOR THE DEGREE OF

MASTER OF TECHNOLOGY

IN

TELEMATICS AND SIGNAL PROCESSING

BY

SHRISHAILAYYA M HIREMATH

Roll No: 208EC110

UNDER THE GUIDANCE OF

Prof. (Dr.) SARAT KUMAR PATRA

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY

ROURKELA, INDIA

2010

Department of Electronics & Communication Engineering

National Institute of Technology, Rourkela

Orissa, India – 769 008

CERTIFICATE

This is to certify that the thesis titled "**ANFIS BASED DATA RATE PREDICTION FOR COGNITIVE RADIO",** submitted by **Shrishailayya M Hiremath,** Roll No. **208EC110 in** partial fulfillment of the requirements for the award of the degree of **Master of Technology** in Electronics & Communication Engineering with specialization in Telematics and Signal Processing , at the National Institute of Technology, Rourkela is an authentic work carried out by him under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University / Institute for the award of any Degree or Diploma.

Date **Prof. Sarat Kumar Patra**

(Supervisor)

HOD, Dept. of E&CE

National Institute of Technology

ROURKELA

Dedicated to all teachers and professors of my two alma maters

SAINIK SCHOOL BIJAPUR and NIT ROURKELA.

"Guru Brahma, Guru Vishnu,

Guru Devo Maheshwara.

Guru Sakshath Parambrahma,

Tasmai Shri Gurave Namaha."

ACKNOWLEDGEMENTS

I am deeply indebted to **Prof. (Dr.) SARAT KUMAR PATRA, HOD**, and Dept. of E&CE, my supervisor on this project, for consistently providing me with the required guidance to help me in the timely and successful completion of this project. In spite of his extremely busy schedules in Department, he was always available to share with me his deep insights, wide knowledge and extensive experience. His advices have value lasting much beyond this project. I consider it a blessing to be associated with him. Also I am very thankful to him for reviewing my thesis.

I would like to express my humble respects to **Prof. G. S. Rath**, **Prof. G. Panda**, **Prof. K. K. Mahapatra, Prof. S. Meher**, **Prof. S. K. Behera, Prof. D.P. Acharya**, and **Prof.A.K.Sahoo** for teaching me and also helping me how to learn. I would like to thanks all faculty members of ECE. Department for their help and guidance. They have been great sources of inspiration to me and I thank them from the bottom of my heart.

I would to like express my special thanks to my research seniors and friends of mobile communication lab, Prasanta, Sanatan, Devirani, Bibhudendra, Bijaya, Sudeendra, Venky, Prgyan and Arun for their help during the research period.

I would like to thank all my friends and especially my classmates for all the thoughtful and mind stimulating discussions we had, which prompted us to think beyond the obvious. I've enjoyed their companionship so much during my stay at NIT, Rourkela.

Last but not least, I would like to thank my parents, who taught me the value of hard work by their own example. They rendered me enormous support being apart during the whole tenure of my stay in NIT Rourkela.

Contents

ACKNOWLEDGEMENTS v
List of Figures viii
List of Tables xi
Acronyms and abbreviations xii
ABSTRACT: xiii
CHAPTER 1 INTRODUCTION
1.1Introduction 1
1.2 Motivation 1
1.3 Objective 2
1.4 Thesis Layout 3
CHAPTER 2 COGNTIVE RADIO
2.1 Introduction 4
2.2 Brief History of CR 4
2.3 Definitions of CR 5
2.4 CR Tasks 6
2.4.1 Radio-scene analysis 7
2.4.2 Channel-State Estimation 9
2.4.3 Predictive Modeling 9
2.4.4 Distributed Transmit-Power Control 10
2.4.5 Dynamic Spectrum Management 10
2.5 Cooperation in Cognitive Radio 13
2.5.1 Virtual Capacity 13
2.5.2 Power Consumption 13
2.5.3 Cost 13
2.5.4 Reliability 14
2.6 Emergent Behaviors of Cognitive Radio 14

2.7 Emerging Cognitive Radio Standards and Deployments 15
2.7.1 IEEE 802.22 15
2.7.2 IEEE 802.11h 15
2.8 Cognitive Radio Applications and Drawbacks........ 16
2.9 Important Institution and forums working on CR 17
CHAPTER 3DATA RATE PREDICTIONS FOR CR USING NEURAL NETWORK SCHEMES
3.1 Introduction 18
3.2 Need For Learning Mechanism 18
3.3 Overview of Different Neural Nets Used for Data Rate Prediction 21
3.3.1 Definition of Neural Network 21
3.3.2 Architecture of neuron model........ 22
3.3.3 Neural Networks Architecture........ 23
3.3.3.1 Feed forward networks architecture and operation: 23
3.3.3.2 Recurrent Neural Networks (RNN)........ 26
3.4 Motivation for Data rate prediction using Neural Network 28
3.5 Basic NN-based Data Rate Prediction........ 29
3.5.1 Preparation procedure........ 29
3.5.2 NN Pattern selection for Basic Scheme 31
3.5.3 Simulation Results and Discussion 32
3.6 Extended NN -based Data Rate Prediction 37
3.6.1 Preparation procedure........ 37
3.6.2 Results and discussion........ 39
3.7 Conclusion........ 45
CHAPTER 4 ANFIS BASED DATA RATE PREDICTION
4.1 Introduction 46
4.2 (ANFIS): An overview........ 46
4.2.1 Basics of Fuzzy Modeling........ 46
4.2.2 ANFIS 49
4.2.3 ANFIS Learning Method........ 52
4.2.4 Membership Function and Rules Selection for ANFIS........ 53
4.3 ANFIS based data rate prediction: Basic Scheme........ 56
4.3.1 Preparation procedure: 56
4.3.2 Simulation results and discussion........ 57
4.4 ANFIS based data rate prediction: Extended Scheme........ 69
4.4.1 Preparation procedure: 69
4.4.2 Simulation results and discussion: 70
4.5 Conclusion........ 74

5.1 Introduction. 75
CHAPTER 5 CONCLUSION
5.2 Contribution of the Thesis.. 75
5.3 Limitations.. 76
5.3 Scope for Future work... 76
Bibliography.. 78
DISSEMINATION OF RESERCH WORK 80

List of Figures

Figure 2. 2 Basic cognitive cycle…………………………………………………………...7

Figure 2. 3 Dynamic spectrum-sharing for OFDM based on four channels………………....12

Figure 3. 1 Simplified representation of cognitive radio cycle………………………………19

Figure 3. 2. Cognitive radio engine……………………………………………………….21

Figure 3. 3 Neuron model………………………………………………………………......23

Figure 3. 4 MLP Architecture……………………………………………………………..25

Figure 3. 5 Elman neural networks……………………………………………………......27

Figure 3. 6 Time window…………………………………………………………………31

Figure 3. 7 Neural network for the basic learning scheme…………………………………...32

Figure 3. 8 Cumulative distribution functions of input time-series………………………….33

Figure 3. 9 MSE curve for -basic scheme...35

Figure 3. 10 Prediction accuracy of selected NN in training sequence-basic scheme……….36

Figure 3.11Prediction accuracy of selected NN in validation sequence-basic scheme………36

Figure 3. 12 Weight values per time slot for the extended scheme………………………….39

Figure 3. 13 Neural network for extended scheme………………………………………..41

Figure 3.14 Prediction accuracy of selected NN in training sequence-Extended scheme..……………………………………………………………………………....42

Figure 3. 15 Prediction accuracy of selected NN in validation sequence-Extended scheme ……………………………………………………………………………………………..42

Fgure 3.16 MSE curve for –extended scheme………………………………………………43

Figure 4. 1 Fuzzy Inference System………………………………………………………48

Figure 4. 2 First order sugeno model…………………………………………………… . .. 51

Figure 4. 3 ANFIS architecture……………………………………………………….....51

Figure 4. 4 Grid partitioning of input space for two input sugeno fuzzy model with nine rules……………………………………………………………………………………..54

Figure 4. 5 ANFIS model for simulation- Basic Scheme……………………………… …...57

Figure 4. 6 Memberships plot for each input before training……………………………….. 59

Figure 4. 7 Membership plot after training……………………………………………… ….59

Figure 4. 8 RMSE for training and validation………………………………………………. 60

Figure 4. 9 Prediction accuracy of conventional ANFIS in training sequence-basic scheme…………………………………………………………………………………..60

Figure 4. 10 Prediction accuracy of conventional ANFIS in validation sequence-basic scheme…………………………………………………………………………………61

Figure 4. 12 Memberships plot for each input after training training in case of FCM based structure………………………………………………………………………… …...62

Figure 4. 13 RMSE for training and validation in case FCM based ANFIS…………………63

Figure 4. 14 Prediction accuracy of FCM based ANFIS in training sequence-basic scheme…………………………………………………………………………………..63

Figure 4. 15 Prediction accuracy of FCM based ANFIS in validation sequence-basic scheme…………………………………………………………………………………..64

Figure 4. 16 Memberships plot for each input before training in case of subtractive clustering based structure……………………………………………………………………………65.

Figure 4. 17 Memberships plot for each input after training in case of subtractive clustering basedstructure…………………………………………………………………………...65

Figure 4. 18 RMSE for training and validation in case subtractive clustering based ANFIS ……………………………………………………………………………………………..66

Figure 4. 19 Prediction accuracy of subtractive clustering based ANFIS in training sequence-basic scheme……………………………………………………………………………….66

Figure 4. 20 Prediction accuracy of subtractive clustering based ANFIS in validation sequence-basic scheme………………………………………………………………… …67

Figure 4. 21 ANFIS model for the extended learning scheme………………………………70

Figure 4. 22 Membership functions before training……………………………………….. ..71

Figure 4. 23 Membership functions after training………………………………………… .72.

Figure 4. 24 RMSE curves after training and validation……………………………………. 72

Figure 4. 25 Prediction accuracy of subtractive clustering based ANFIS in training sequence-extended scheme scheme……………………………………………………………… .73

Figure 4. 26 Prediction accuracy of subtractive clustering based ANFIS in validation sequence-extended scheme…………………………………………………………...73

List of Tables

Table 3.1: Performance index for different NN.-basic scheme…………………………… .37

Table 3.2: Performance index of NNs - extended case…………………………………..44

Table 4.1: Performance index of all ANFIS techniques-basic scheme…………………….68

Table 4.2: Performance index of ANFIS and NNs techniques -extended case……………74

Acronyms and abbreviations

ANFIS	Adaptive neuro-fuzzy inference system
AI	Artificial Intelligence
ANN or NN	Artificial neural networks
CR	Cognitive radios
CSI	Channel-state information
DARPA	Defense Advanced Research Projects Agency
DAB	Digital audio broadcast
DVB	Digital video broadcast
FCC	Federal Communications Commission
FF	Feed-forward networks
FCM	Fuzzy C-means clustering
MSE	Mean square error
MLP	Multilayer perceptron
RAT	Radio access technology
RNN	Recurrent Neural Networks
RMSE	Root mean square error
SC	Subtractive Clustering
TRAI	Telecom Regulation Authority of India

ABSTRACT

Intelligence is needed to keep up with the rapid evolution of wireless communications, especially in terms of managing and allocating the scarce, radio spectrum in the highly varying and disparate modern environments. Cognitive radio systems promise to handle this situation by utilizing intelligent software packages that enrich their transceiver with radio-awareness, adaptability and capability to learn. A cognitive radio system participates in a continuous process, the ‘‘cognition cycle”, during which it adjusts its operating parameters, observes the results and, eventually takes actions, that is to say, decides to operate in a specific radio configuration (i.e., radio access technology, carrier frequency, modulation type, etc.) expecting to move the radio toward some optimized operational state. In such a process, learning mechanisms utilize information from measurements sensed from the environment, gathered experience and stored knowledge and guide in decision making. This thesis introduces and evaluates learning schemes that are based on adaptive neuro-fuzzy inference system (ANFIS) for predicting the capabilities (e.g. data rate) that can be achieved by a specific radio configuration in cognitive radio.

First a ANFIS based scheme is proposed. The work reported here is compare previous neural network based learning schemes. Cognitive radio is a intelligent emergent technology, where learning schemes are needed to assist in its functioning. ANFIS based scheme is one of the good learning Artificial intelligence method, that combines best features of neural network and fuzzy logic. Here ANFIS and neural networks methods are able to assist a cognitive radio system to help in selecting the best one radio configuration to operate in. Performance metric like RMSE, prediction accuracy of ANFIS learning has been used as performance index.

CHAPTER 1

Introduction

1.1 Introduction

An electromagnetic radio spectrum is a natural resource; the use of it is licensed by government agencies. With entry of advanced high data rate wireless applications and existing wireless system upgrading has lead to scarcity in spectrum. Currently, spectrum allotment operates by providing each new service with its own fixed frequency block. As day passes demand for spectrum increasing rapidly and it would be increasing in future. Since the most technologies are moving towards fully wireless. Unlicensed new technologies like Digital video broadcast (DVB), Digital audio broadcast (DAB), internet, Wimax etc. launched recently and are reaching thousands of customers in rapid speed. Most of the primary spectrum is assigned, so it is becoming very difficult to find spectrum for either new services or expanding existing infrastructure. Present government policies do not allow unlicensed access of licensed spectrum, constraining them instead to several heavily populated, interference-prone frequency bands. As the result there is huge spectrum scarcity problem.. In spite of high licensing fees, this resource is presently underutilized. In particular, if we were to scan the radio spectrum, including the revenue-rich urban areas, we would find that some frequency bands in the spectrum are unoccupied for some of the time, and many frequency bands are only partially occupied, whereas the remaining frequency bands are heavily used [**1**] . It is therefore not surprising to find that underutilization of the radio spectrum is being challenged on many fronts, including Federal Communications Commission (FCC) in the United States, Telecom Regulation Authority of India (TRAI)of India.

1.2 Motivation

Cognitive radio offers a novel way to solving spectrum underutilization problems. It does this by sensing the radio environment with a twofold objective: (I) identifying those sub bands of the radio spectrum that are underutilized by the primary (i.e., legacy) users and (ii) providing the means for making those bands available for use by unserviced secondary users.

The idea of cognitive radio goes beyond making productive use of unused part of the spectrum and being capable of making human-like decisions to transmit without obstruction. With an aim to achieve above ability, cognitive radio behave in reactive or proactive manner based on external environmental information's, as well as their goals, principles, capabilities, experience and knowledge. In this regard future radio need to be intelligent enough in selecting radio configuration, by taking in account of device operation status and environmental conditions, goals, policies, profiles and machine learning. In this respect, [**2**] [**3**] [**4**]future cognitive radio devices will have the capability, or luxury, to choose on the fly the radio configuration, by taking into account the context of operation (device status and environment aspects), goals, policies, profiles and capabilities, and machine learning (for representing and managing knowledge and experience). In the more general sense, the term radio configuration or simply configuration refers to a chosen carrier frequency and a specific radio access technology (RAT) but can be extended to include other operating parameters like transmit power, modulation type, etc.

This definition also allows a spectrum band to be used for operating in different RATs, in accordance with the flexible spectrum management concept [**4**].Cognitive radio involves different stages which are explained in later chapter. Objective of work is to bring in the learning capability in channel estimation and predictive modeling phase, for improving the stability and reliability of the discovery and evaluation of the configuration capabilities, without relying solely on the recent measurements. To this effect, many different learning techniques are available and can be used by a cognitive radio ranging from pure lookup tables to arbitrary combinations of machine learning techniques that include artificial neural networks, ANFIS, evolutionary/ genetic algorithms, reinforcement learning, hidden Markov models, etc. In this work various neural networks based data rate prediction based on previous work have been tested and an ANFIS model based technique for data rate prediction in assisting cognitive radio has been proposed.

1.3 Objective

Main of aim of work is to explore Neural network and ANFIS based learning scheme for cognitive radio and aim at solving the problem related to the channel estimation and predictive modeling phase of cognitive radio systems and stated as follows: ''Given a candidate radio configuration, what are its anticipated capabilities (e.g., in terms of achievable data rate), taking into account recent information sensed, as well as the past experience and knowledge?".Here learning schemes are considered for basic and extended based on neural networks and ANFIS are designed to enhance the learning capabilities of a cognitive terminal, in terms of assisting it to predict the data rate that a specific radio configuration could achieve if it was selected for operation.

1.4 Thesis Layout

The thesis is organized as follows. Chapter 2 gives an introduction to cognitive radio. Here we discuss various stages of cognitive radio, its emergent behavior and standards also applications and various research organization dealing with CR. Chapter 3 discusses neural networks for data rate prediction. Here basic overview of different NNs used for Data rate perdition and Need for NNs to assist CR, also procedure and model for data rate prediction are explained considering basic and extended scheme. Chapter 4 proposes ANFIS based learning scheme and its use for data rate prediction for basic scheme and extended scheme. Performance of ANFIS methods is also compared with existing neural network methods. Chapter 5 discusses conclusion and future work to be done.

CHAPTER 2

COGNITIVE RADIO

2.1 Introduction

As we know wireless field is rapidly evolving. It all started with Marconi first radio broadcast from England to one person in Nova Scotia. It was the radio equivalent of flying an airline's passengers across the Atlantic one at a time. [**5**]Today, we pack radio waves into the air as tightly as economy seats on a 747. New radio technologies keep coming along Wi-Fi, WiMax, Bluetooth, ZigBee, the growing panoply of cellular voice and digital services, broadcast satellite, and more. Due to the large number of standards, spectrum availability has become an important issue. Spectrum usage regulations prohibit unlicensed users to operate in a licensed spectrum. However, it has been observed that not the entire licensed spectrum is used at all places all the time. An unlicensed user can take advantage of such a situation to communicate thereby increasing spectrum efficiency. This is the basic idea behind Cognitive Radio (CR) [**1**].This chapter provides a brief history, definition, functionality, applications of CR and List institutes and forums working on it.

2.2 Brief History of CR

The Cognitive radios(CR) do not have the history of a century; rather the development of cognitive radio is still at a research stage. The Cognitive Radio is an emerging technology, for the efficient use of the limited available spectrum. Nevertheless, as we look to the future, we see that cognitive radio has the capacity to make a significant difference to the way the radio spectrum can be accessed, with much improved utilization. Indeed, given its potential, the cognitive radio can justifiably be described as a "disruptive, but unobtrusive technology". Disruptive as it can make a great difference in the [**6**]technology. Unobtrusive as it attracts with its solution for utilization of the already licensed frequency bands efficiently. It all started when [**3**] Joseph Mitola III coined word "Cognitive Radio" in his doctoral dissertation. He described, the way a cognitive radio could enhance the flexibility of personal wireless services, through a new language called the "Radio Knowledge Representation Language" (RKRL) [**7**]. The idea of RKRL was further expanded in Mitola's doctoral dissertation, presented at the Royal Institute of Technology, Sweden, in May 2000 [**3**]. This dissertation presents a conceptual overview of CR as an exciting multidisciplinary subject.

Cognitive radio is not a single technology to be very explicit. It resulted from many technologies coming together to result in the Cognitive Radio technologies, due the verity that exists among its applications. For example, the development of digital signal processing (DSP), development of math and signal processing tools and source coding of voice, image and data etc..It is technology which is built on Software defined Radio which is brain child of Defense Advanced Research Projects Agency (DARPA). DARPA later launched the neXt Generation program (XG) that focused on "the enabling technologies and system concepts to dynamically redistribute allocated spectrum." After the successful conclusion of the XG project, the FCC realized that CR was the answer to stimulate growth of open spectrum [**8**].

2.3 Definitions of CR

After Mitola coined the word "Cognitive radio" its definition is also evolving as research interest in CR is increasing. Regulatory bodies, prominent researchers and forums define it in different ways.

According to Mitola [**3**], CR is defined as ".*The point in which wireless personal digital assistants (PDAs) and the related networks are sufficiently computationally intelligent about radio resources and related computer-to-computer communications to: (a) detect user communications needs as a function of use context, and (b) to provide radio resources and wireless services most appropriate to those needs".* However the concept of CR is not limited strictly to wireless devices such as PDAs

Widely cited paper by Simon Haykin on CR defines it as follows [**1**]: "*Cognitive radio is an intelligent wireless communication system that is aware of its surrounding environment (i.e., outside world), and uses the methodology of understanding-by-building to learn from the environment and adapt its internal states to statistical variations in the* *incoming RF stimuli by making corresponding changes in certain operating parameters (e.g., transmit-power, carrier-frequency, and modulation strategy) in real-time, with two primary objectives in mind:*

- *Highly reliable communications whenever and wherever needed;*
- *Efficient utilization of the radio spectrum.*

The regulatory bodies which focus on the operation of transmitter like FCC defines cognitive radio as: *"A radio that can change its transmitter parameters based on interaction with the environment in which it operates."* [**9**] [**10**].

While aiding the FCC in its efforts to define cognitive radio, IEEE USA offered the following definition [**10**] : *"A radio frequency transmitter/receiver that is designed to intelligently detect whether a particular segment of the radio spectrum is currently in use, and to jump into (and out of, as necessary) the temporarily-unused spectrum very rapidly, without interfering with the transmissions of other authorized users."*

SDR forum which is one of the highly associated with CR and SDR , that works on CR [**10**]application defines CR as *"An adaptive, multi-dimensionally aware, autonomous radio (system) that learns from its experiences to reason, plan, and decide future actions to meet user needs."*

So among all definition it is found that find following terminologies are common "**Observation", "Adaptability" and "Intelligence".** Using following terminologies CR is defined as [**10**] :

"Radio whose control processes permit the radio to leverage situational knowledge and intelligent processing to autonomously adapt towards some goal.

2.4 CR Tasks

In [**1**], a typical cognitive radio operation is presented as a simplification to the ‘‘cognition cycle" initially described in [**3**] [**7**]and can be divided into three, tightly interconnected tasks (see Fig. 1).

1) **Radio-scene analysis, which encompasses the following**:

• estimation of interference temperature of the radio environment;

• detection of spectrum holes.

2) Channel identification, which encompasses the following:

• estimation of channel-state information (CSI);

• prediction of channel capacity for use by the transmitter.

3) Transmit-power control and dynamic spectrum management.

Tasks 1) and 2) are carried out in the receiver, and task 3) is carried out in the transmitter. Through interaction with the RF environment, these three tasks form a cognitive cycle, which is presented in its most basic form in Figure.2.1.

2.4.1 Radio-scene analysis

During Radio-scene analysis different radio configurations are probed to estimate interference temperature of the radio environment and to detect of spectrum holes. Interference temperature and spectrum holes are discussed.

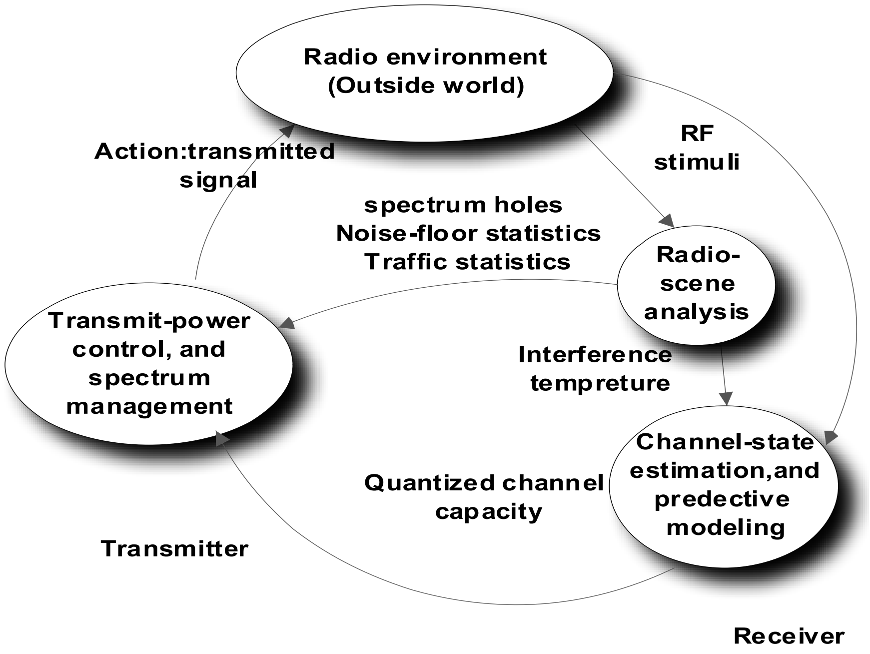

Figure 2. 1 Basic cognitive cycle.

Interference Temperature: [**11**]

The interference temperature is a measure of the sensed power in a certain frequency band. Thus, by obtaining this measure, two important limits can be identified:

• The maximum level where any signal exceeds threshold level.

• The minimum level where any signal below it can be neglected and thus that certain band can be considered as empty or unused, and can be used by other users.

Spectrum Hole: [**1**]

A spectrum hole is a band of frequencies assigned to a primary user, but, at a particular time and specific geographic location, the band is not being utilized by that user. Primary users are those holding licensed channels or primary bands.

As said above radio scene analysis includes two functionality. These two stages are performed periodically. The interference temperature is suggested to be estimated for the whole targeted frequency ranges. Then depending on the current interference and the interference temperature on the previous iterations all channels can be classified into three types of spectrum holes:

• White spectrum holes, which are fully not used.

• Gray spectrum holes, which are partially used.

• Black spectrum holes, which are fully used.

After the sensing operation completed, the users will be allowed to freely use the white holes and partially use the gray holes in a way that will not disturb the primary user. But they will not use the black holes, because the black holes are assumed to be fully used and any extra use will interfere with the ongoing communication on them. In general, there are two sensing modes, reactive sensing and proactive sensing, depending on the way to initiate the sensing. These two modes can be defined as follows:

Reactive sensing: The sensing is initiated only when the user has data to send, thus it is called on-demand sensing. If no usable channel was found, the user will wait for a predefined time and then restart sensing again until the user send all data that he was trying to send. This technique reduces the sensing overhead. And its Disadvantage is that, data is delayed until the sensing is performed with a good accuracy.

Proactive sensing: The sensing is done periodically even when the user is not intending to send any data. The time between the sensing iteration is called the sensing period. These sensing periods may differ between the channels since each channel has its own unique behavior. The sensing periods should be optimized separately for each channel to compensate for the unique traffic pattern on that channel.

Advantage: The delay is decreased since the users will know the holes even before they need them.

Disadvantage: A lot of time and effort is wasted on sensing even when it is not needed, thus increasing the sensing overhead.

Each one of those two modes has its advantages and its disadvantages, thus both of them might be used depending on the application and the environmental conditions.

2.4.2 Channel-State Estimation [1] [11]

Channel estimation was also proposed to be part of the cognitive radio. This operation aims in analyzing the channel behavior and its effects on the transmitted signal and estimating the impulse response of the channel. By knowing the channel impulse response, information neutralized by the receiver for an equalizer design or on the transmitter by transmitting a signal that can absorb those effects.

2.4.3 Predictive Modeling [11]

Due to the dynamic behavior of the communication channels and environment, analyzing only the current channel and using the results directly to select a free channel or to equalize the channel might leads to an inefficient use of the resources. Thus, developing a more accurate way to use the knowledge obtained from the normal analysis results is the predictive modeling. It aims on finding models that predicts the behavior of the channels on the future and even the traffic patterns. Using those models will increase the efficiency of using the analysis results and will improve and ease the decision taking procedures. The predictive modeling uses the current observations along with the previous observations and based on some statistical measures it tries to find the model that will most likely suits the channel or the traffic in the near future. Usually, prediction implies the possibility for some errors. But it significantly improves the performance of the system to an extent where those errors can be neglected. Predictive modeling stage, where thesis analysis involves. This stage is explained later.

2.4.4 Distributed Transmit-Power Control [1]:

Like the spectrum allocation, this process is done centrally in conventional radios. Thus, in cognitive radio each user should take care of it is own transmission power control and gives some feedbacks regarding the signals that it received. As a result, the power control process will be done in a distributed manner. In other words, each user must make sure that the signal that he transmits will reach the receiver in a certain level high enough to be detected by the receiver and low enough to avoid interfering with other users. In the same time each user has to inform the users, which are transmitting to it, about the reception signal level. The power control operation plays a crucial part in minimizing the interference and in insuring the needed quality of service in many communication systems.

2.4.5 Dynamic Spectrum Management [1]

As with transmit-power control, dynamic spectrum management (also referred to as dynamic frequency-allocation) is performed in the transmitter. These two tasks are intimately related to each other, and hence have been included inside a single functional module, which performs the role of *multiple-access control* in the basic cognitive cycle of Figure. 2.1 simply replace, the primary purpose of spectrum management is to develop an adaptive strategy for the efficient and effective utilization of the RF spectrum. Specifically, the *spectrum-management algorithm* is designed to do the following. *Building on the spectrum holes detected by the radio-scene analyzer and the output of transmit-power controller, select a modulation strategy that adapts to the time-varying conditions of the radio environment, all the time assuring reliable communication across the channel.* Communication reliability is assured by choosing the SNR gap large enough as a design parameter.

Modulation Considerations

A modulation strategy that commends itself to cognitive radio is the OFDM by virtue of its flexibility and computational efficiency. For its operation, OFDM uses a set of carrier frequencies centered on a corresponding set of narrow channel bandwidths. Most important, the availability of rate feedback (through the use of a feedback channel) permits the use of *bit loading*, whereby the number of bits/symbol for each channel is optimized for the SNR characterizing that channel; As time evolves and spectrum holes come and go, the bandwidth-carrier frequency implementation of OFDM is dynamically modified, as illustrated in the time-frequency picture in Figure.2.2 for the case of four carrier frequencies. The picture illustrated in Figure.2.2 describes a distinctive feature of cognitive radio: a *dynamic spectrum-sharing process*, which evolves in time.

In effect, the spectrum-sharing process satisfies the constraint imposed on cognitive radio by the availability of spectrum holes at a particular geographic location and their possible variability with time. Throughout the spectrum-sharing process, the transmit-power controller keeps an account of the bit-loading across the spectrum holes in use. In effect, the dynamic spectrum manager and the transmit-power controller work in concert together, thereby fulfilling the multiple-access control requirement. Starting with a set of spectrum holes, it is possible for the dynamic spectrum management algorithm to confront a situation where the prescribed FER cannot be satisfied. In situations of this kind, the algorithm can do one of two things:

i) work with a more spectrally efficient modulation strategy, or else;

ii) incorporate the use of another spectrum hole, assuming availability.

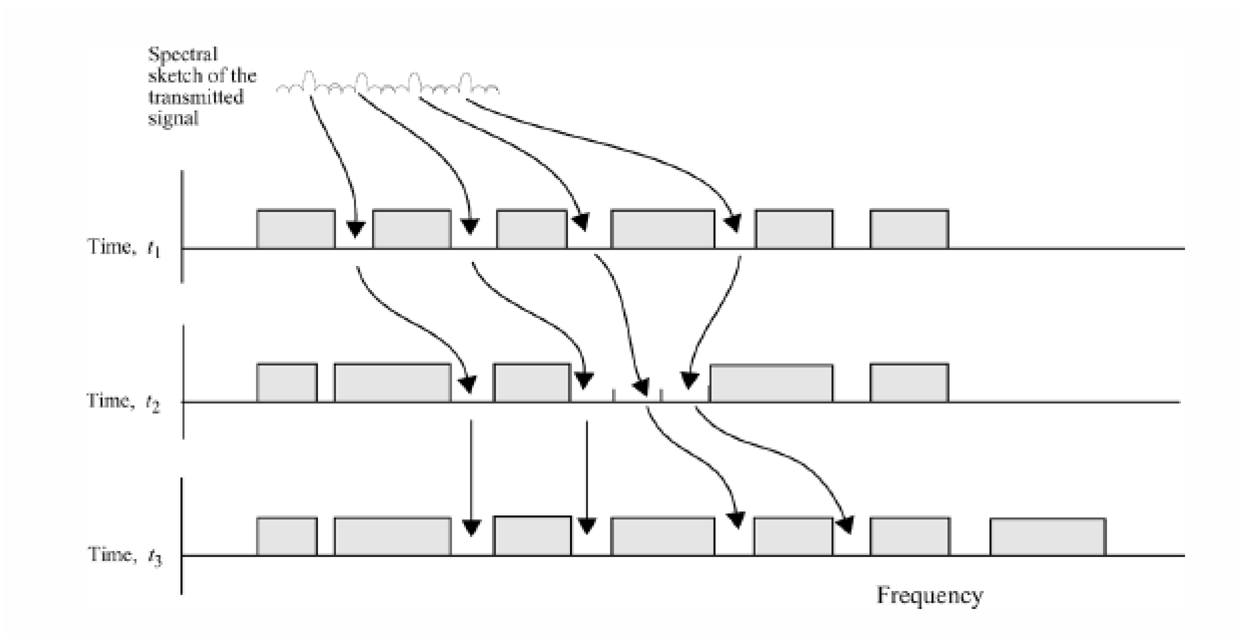

Figure 2. 2 Dynamic spectrum-sharing for OFDM based on four channels [1]

In approach i), the algorithm resorts to increased computational complexity, and in approach ii), it resorts to increased channel bandwidth so as to maintain communication reliability.

Traffic Considerations [1]

In a code-division multiple-access (CDMA) system, like the IS-95, there is a phenomenon called *cell breathing*: the cells in the system effectively shrink and grow over time. Specifically, if a cell has more users, then the interference level tends to increase, which is counteracted by allocating a new incoming user to another cell; that is, the cell coverage is shrunk. If, on the other hand, a cell has less users, then the interference level is correspondingly lowered, in which case the cell coverage is allowed to grow by accommodating new users. So in a CDMA system, the traffic and interference levels are associated together. In a cognitive radio system, based on CDMA, the dynamic spectrum management algorithm naturally focuses on the allocation of users, first to white spaces with low interference levels, and then to grey spaces with higher interference levels. When using other multiple-access techniques, such as OFDM, co-channel interference must be avoided. To satisfy this requirement, the dynamic-spectrum management algorithm must include a *traffic model* of the primary user occupying a black space. The traffic model, built on historical data, provides the means for predicting the future traffic patterns in that space. This in turn, makes it possible to predict the duration for which the spectrum hole vacated by the incumbent primary user is likely to be available for use by a cognitive radio operator. In a wireless environment, two classes of traffic data patterns are distinguished, as summarized here.

1) Deterministic patterns. In this class of traffic data, the primary user (e.g., TV transmitter, radar transmitter) is assigned a fixed time slot for transmission. When it is switched OFF, the frequency band is vacated and can, therefore, be used by a cognitive radio operator.

2) *Stochastic patterns*. In this second class, the traffic data can only be described in statistical terms. Typically, the *arrival times* of data packets are

modeled as a *Poisson process;* while the service times are modeled as *exponentially distributed*, depending on whether the data are of packet-switched or circuit-switched kind, respectively. In any event, the model parameters of stochastic traffic data vary slowly and, therefore, lend themselves to on-line estimation using historical data. Moreover, by building a *tracking strategy* into design of the predictive model, the accuracy of the model can be further improved.

2.5 Cooperation in Cognitive Radio [1] [11]

Cooperation is another novel approach in cognitive radio. It aims in improving the performance of the users who lack some of the communication resources by the help of the users who can easily access those resources. This improves the efficiency of the resources usage as well as the users' quality of service. By realizing this approach many benefits and services are expected to emerge. In the following sections, some of these benefits are introduced

2.5.1 Virtual Capacity

Cooperation between the cognitive radios is expected to add some sort of virtual capacity, which is the capacity that gains by indirect links. In contrast of the long range links that have usually small capacity, the short range links can be said to have very high capacity. Thus by establishing a number of short range links with its neighbors, a user can use all of his neighbors low capacity links. Thus, the user can virtually form a high capacity link to a destination it individually cannot establish a link with such capacity. Optimizing this operation will lead to extreme increase in the capacity of the users and improve the overall spectral efficiency.

2.5.2 Power Consumption

By cooperating, users can avoid using long range links and use short range links. Since long links consumes much more energy than the short ones, this will dramatically decrease the overall system power consumption.

2.5.3 Cost

Since all cooperative links are not managed by a service provider, they are free. Thus from users' point of view, it is preferable if they can use such links instead of the expensive and low capacity links. Moreover, from the service providers' point of view, their system will have more traffic. In addition, without the need for increasing the physical range of their sites, their virtual coverage area will increase, and the cost of the devices will decrease since the power consumption will be low and thus the need for costly and heavy duty batteries will be minimized.

2.5.4 Reliability

In contrast of the conventional single link communication, using cooperative multilink communication offers higher reliability. This is caused by the fact that the effect of losing one link from the multilink set is by far smaller than the total lost in the case of losing a link in the case single link communication. Moreover, the probability of losing all of the links in a multilink communication is extremely low compared to the probability of loosing a single link.

2.6 Emergent Behaviors of Cognitive Radio [1]

Considering the complex situations that should normally happens in a cognitive radio system, the rapidly varying configurations of the radios my lead the system to a new mysterious state. Since radios may compete or cooperate in cognitive radio system, their benefits can normally oppose each other. Thus there is no simple way to ensure agreeing on an acceptable solution for all sides. The states can be categorized into two main types:

Positive state: where the overall system performance is improving and the radio resources are used efficiently.

Negative state: where the overall system performance is degrading and some of the radio resources are not used or are used inefficiently.

Thus, it is very important to develop a way to ensure the system convergence to a positive state and avoid approaching a negative state. This can be done by developing models that can foresee the system development. This kind of a system can be handled by one of these two models, self organizing system or evolutionary game. The self organizing system model views the system as a group of players, each player actions are influenced by the others and therefore every action is reflected back to its originator. If these reflections continue in a pattern that amplifies their effects they may lead to instability in the system. Thus, the system reflections and their effects need to be checked in order to protect the system from the negative states.

The evolutionary game model views the system as a group of animals every one of them has its own instincts and intentions which lead to behaviors that looks very stochastic. Using those two models the system behaviors can be predicted. From these predictions, further investigations can be made to find the parameters that will likely lead to positive state and use them and find the ones that will probably lead to negative state and avoid them. It is also possible to find simple rules that can ensure the system stability.

2.7 Emerging Cognitive Radio Standards and Deployments [10] [12]

The IEEE 802 community is currently developing two standards that directly relate to cognitive radio – IEEE 802.22 also called standard for cognitive wireless regional area networks (WRANs) and 802.11h. Additionally, 802.11k is developing techniques for incorporating radio resource management information into WLAN operation – in effect incorporating knowledge about the environment and the radios.

2.7.1 IEEE 802.22

There are three applications typically discussed for coexistence with initial trial deployments of cognitive radios: television, microwave point-to-point links, and land mobile radio. Each of these applications has been shown to dramatically underutilize spectrum on average. However, only television signals have the advantage of incumbent signals that are easy to detect (as opposed to a microwave point-to-point links) and not involved in life-critical applications (as would be the case for many land mobile radio systems). Throughout its history, the UHF bands were under-allocated as regulators underestimated the cost-effectiveness of establishing new TV towers in these bands. It was not until the advent of cable TV that smaller TV stations were capable of cost-effective operation. Now with the introduction of HDTV technology, regulators in the US plan to force a nation-wide switch to this more efficient modulation by 2009 [**12**] accompanied by a completion of a de-allocation from analog TV of 108 MHz of high quality spectrum. With these bands in mind, the 802.22 working group is pursuing the development of a waveform intended to provide high bandwidth access in rural areas using cognitive radio techniques. In a report presented at DySPAN [**12**], it is stated that the 802.22 standard intends to achieve spectral efficiencies of up to 3 bits/sec/Hz corresponding to peak download rates at coverage edge at 1.5 Mbps. Simultaneously, the 802.22 system hopes to achieve up to 100 km in coverage.

2.7.2 IEEE 802.11h

Unlike 802.22, 802.11h is not formulated as a cognitive radio standard. However, the World Wireless Research Forum has noted that a key portion of the 802.11h protocol –

dynamic frequency selection – has been termed a “cognitive function”. Reasons for 802.11h WLAN might be considered as CR based on following tasks.

Observation – It requires WLANs to estimate channel characteristics such as path loss and link margin and further requires the radios estimate channel characteristics such as path loss and link margin.

Orientation – Based on these observations, the WLAN has to determine if it is operating in the presence of a radar installation, in a bad channel, in band with satellites, or in the presence of other WLANs.

Decision – Based on the situation that the WLAN is encountering, the WLAN has to decide to change its frequency of operation (*Dynamic Frequency Selection*), adjust the transmit power (*Transmit Power Control*), or both.

Action – The WLAN has to then implement this decision.

Reviewing most of the definitions from before, only learning or “recalling and correlating past actions, environments and performance” is not required as part of the standard. However, if we move beyond the requirements of the standard to expected implementations, it seems reasonable that many vendors will include and leverage some memory of past observations (useful for detecting intermittent transmitters) which implies that both cognitive radio definitions will be satisfied.

2.8 Cognitive Radio Applications and Drawbacks

Some of the important applications of CR are as follows:

- Improving spectrum utilization & efficiency
- Improving link reliability
- Less expensive radios
- Advanced network topologies
- Enhancing SDR techniques
- Automated radio resource management.

Drawbacks of CR.

- Security
- Software reliability
- Keeping up with higher data rates
- Loss of control

- Regulatory concerns
- Significant research remains to be done to realize commercially practical cognitive radio

2.9 Important Institution and forums working on CR [8] [10]

Some of the important institution, forums and research organization where extensive research is going on are listed below.

DARPA (USA)- is exploring many different aspects of cognitive radio as part of the xG program and other ongoing programs. Unfortunately, many of the results of the DARPA programs
are not currently in the public domain

IEEE- has started the IEEE 1900 group to study the issue of cognitive radio and giving standard like 802.22

SDR [**13**] Forum- chartered two groups in 2004 to explore cognitive radio issues: the Cognitive Radio Working Group and the Cognitive Radio Special Interest Group. The working group is tasked with standardizing a definition of cognitive radio and identifying the enabling technologies for cognitive radio.

FCC(USA)- On May 19, 2003, the FCC convened a workshop to examine the impact that cognitive radio could have on spectrum utilization and to study the practical regulatory issues that cognitive radio would raise.

Virginia Tech(USA), Work is being performed exploring techniques to exploit collaborative radio to improve network performance.

Win lab [**14**] Rutgers University is developing a cognitive radio test bed for disaster response using commercially available components

E ^{2} R- European initiative into supporting End-to-End Reconfigurability with numerous participating European universities and companies

BWRC [**15**] is currently developing a cognitive radio for sensing and opportunistically using the spectrum.

CHAPTER 3

DATA RATE PREDCITION FOR CR USING NEURL NETWORKS SCHEMES

3.1 Introduction

As defined in previous chapter CR is adaptive radio which try to change its behavior according to environmental conditions and try to adjust its parameters to provide efficient utilization of spectrum and make radio more intelligent. So in this regard intelligence needs to embed in the cognitive network. As we discussed before there are three stages in cognition cycle, in any of the stages intelligence can be put in. Intelligence is nothing but learning mechanisms that are capable of exploiting measurements sensed from the environment, gathered experience and stored knowledge, are judged as rather beneficial for guiding decisions and actions. This work presents learning mechanism that is introduced in channel estimation and predictive module phase to predict data rate of particular radio configuration which is previously done in [**4**].

In this Chapter first section discuses need for learning mechanism for CR, second section gives the overview of different neural networks used for prediction, fourth section discusses motivation for data rate prediction using NNs ,fifth section gives algorithm used for data rate prediction in basic scheme and simulation results for the same, sixth section disuses data rate prediction algorithm for extended case and simulation results and last section gives the conclusion.

3.2 Need For Learning Mechanism [4].

To achieve better spectrum efficiency radio device need to be equipped with cognitive capabilities discussed in previous chapter. From previous chapter it is understood that, cognitive systems determine their behavior in a reactive or proactive manner, based on the external, environmental stimuli, as well as their goals, principles, capabilities, experience and knowledge. In this respect [**4**], future cognitive radio devices will have the capability, or luxury, to choose on the fly the radio configuration, by taking into account the context of operation (device status and environment aspects), goals, policies, profiles and capabilities, and machine learning (for representing and managing knowledge and experience). Simplify the working of previous cognition cycle we define term called radio configuration. It refers to a chosen carrier frequency and a specific radio access technology (RAT) but can be extended to include other operating parameters like transmit power, modulation type, etc. This definition also allows a spectrum band to be used for operating in different RATs, in accordance with the flexible spectrum management concept. Previous chapter cognition cycle is modified in this chapter define problem statement. It is shown as tightly interconnected three phase system in Figure 3.1.

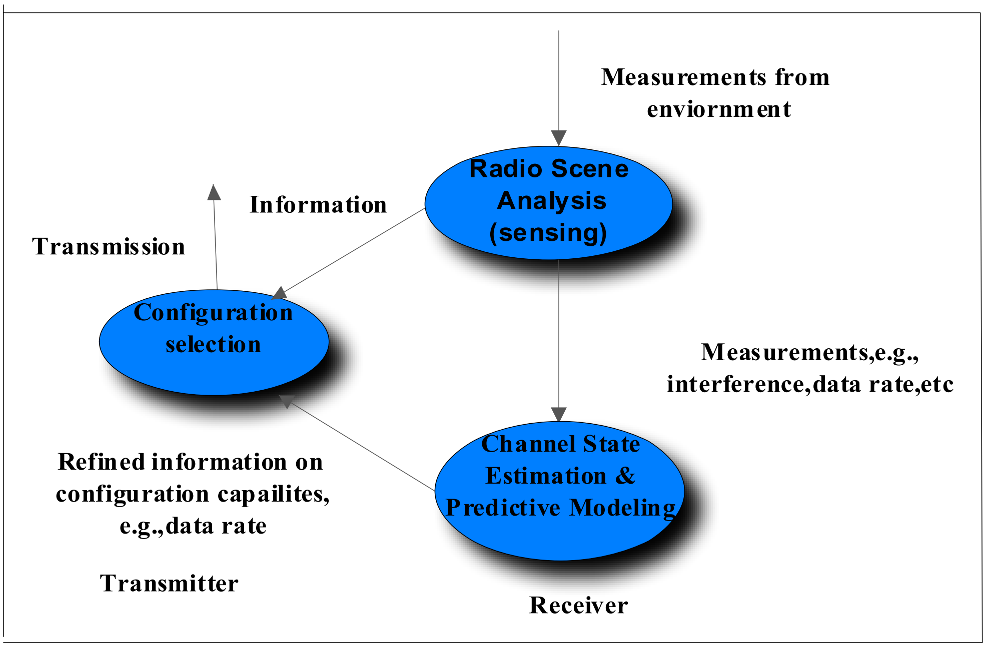

Figure 3. 1 Simplified representation of cognitive radio cycle.

Radio-scene analysis: during which different configurations are probed, and the respective environment conditions, e.g. interference related, are sensed.

Channel estimation and predictive modeling: during which the capabilities of configurations are discovered (discovery process) and accordingly assessed, based on the measurements of the previous phase; moreover, past experience and knowledge can be exploited in this phase.

Configuration selection: during which the transmitter sends the desired signal by means of the ‘‘best” radio configuration (RAT, frequency, modulation, transmit power, etc.), as it derives from the information of the previous two phases.

The approach that is adopted here is that a cognitive radio results from the enhancement of a software radio with cognitive capabilities. Those capabilities are often provided by an intelligent software package called a cognitive engine, as mentioned previously and depicted in Figure.3.2, albeit, according to FCC ,'neither having software nor being programmable are requirements of a cognitive radio'.

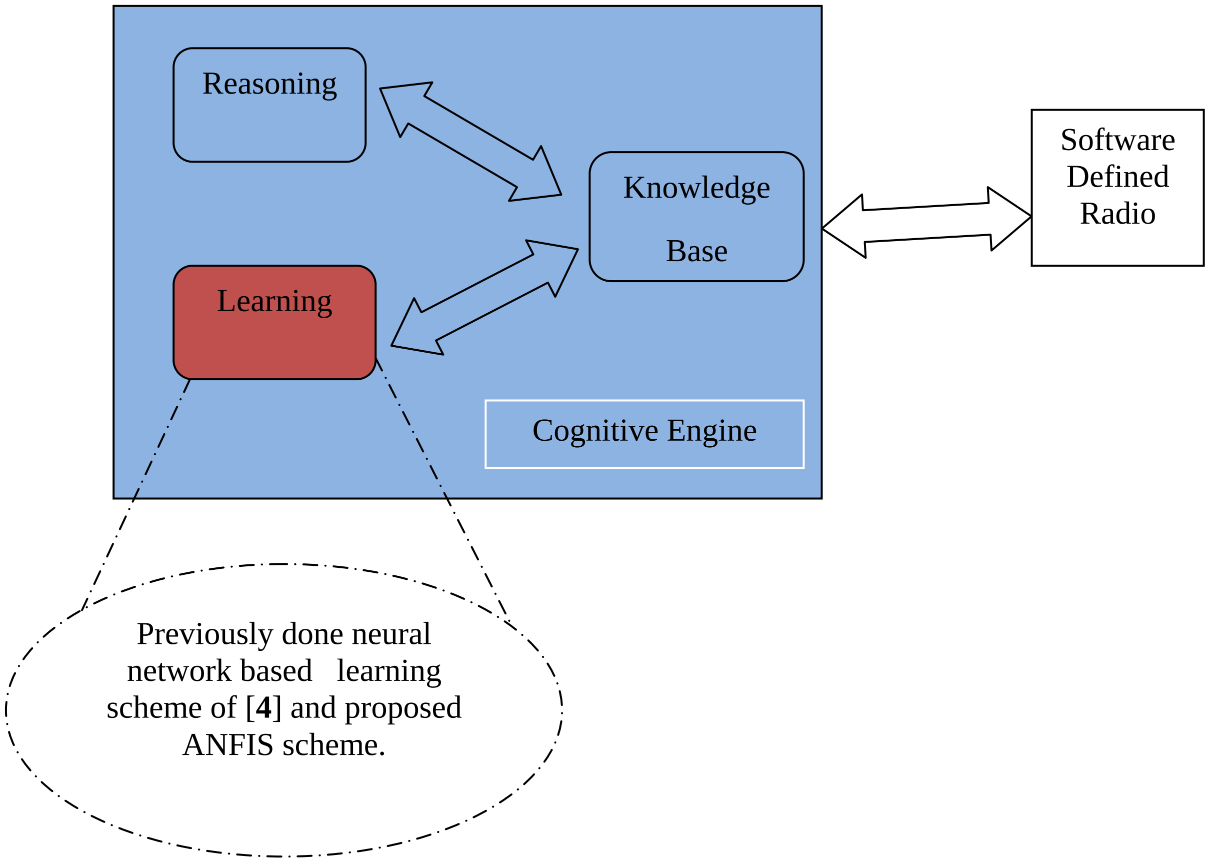

Figure 3. 2. Cognitive radio engine.

As said earlier within the cognition cycle, the cognitive engine derives and enforces decisions to the software-based radio by continuously adjusting its parameters, observing and measuring the outcomes and taking actions to move the radio toward some desired operational state. Meanwhile, cognitive radios are capable of learning lessons and storing them into a knowledge base, from where they may be retrieved, when needed, to assist future decisions and actions. A reasoning engine determines which actions are executable in each environment state. Considering that this could result in a computationally intensive and time-consuming process, e.g. in case of a diversified radio environment with numerous state-action pairs, the need for a learning engine seems to be imperative. The integration of a learning engine can be important especially for the channel estimation and predictive modelling phase, for improving the stability and reliability of the discovery and evaluation of the configuration capabilities, without relying solely on the recent measurements. To this effect, many different learning techniques are available and can be used by a cognitive radio ranging from pure lookup tables to arbitrary combinations of machine learning techniques that include artificial neural networks, ANFIS, evolutionary/ genetic algorithms, reinforcement learning, hidden Markov models, etc. This chapter discusses neural network based learning scheme, that aims at solving the problem related to the channel estimation and predictive modeling phase of cognitive radio systems of paper [**4**].This work is used to compare the ANFIS based results which are discussed in next chapter as the thesis contribution.

Problem statement: Given a candidate radio configuration, what are its anticipated capabilities (e.g., in terms of achievable data rate), taking into account recent information sensed, as well as the past experience and knowledge?"

Main objective is study two learning schemes, a 'basic' and an 'extended' one, that are based on neural networks and are designed to enhance the learning capabilities of a cognitive terminal, in terms of assisting it to predict the data rate that a specific radio configuration could achieve if it was selected for operation, and at last to give benchmarking on neural networks .This Neural networks schemes are used to compare final thesis contribution that is learning based on ANFIS.

3.3 Overview of Different Neural Nets Used for Data Rate Prediction

3.3.1 Definition of Neural Network

Neural nets are the part of Artificial Intelligence (AI).Work on neural net has been motivated right from its inception by the recognition that human brain computes in an entirely different way from the conventional digital computer. Human brain is considered as the highly *complex*, *nonlinear,* and *parallel computer* (information processing system) [**16**]. Human brain can perform tasks much faster than the fastest existing computer thanks to its special ability in massive parallel data processing. This functionality is due to biological neural network, which are connected physically to the human nervous system located in a human brain. Artificial neural networks (ANN or simply NN), are made up of artificial neurons interconnected to each other to form a programming structure that mimics the behavior and neural processing (organization and learning) of biological neurons. NNs try to mimic such a providential behavior for solving narrowly defined problems i.e., problems with an associative or cognitive tinge. To this effect, NNs have been extensively and successfully applied to [**17**] pattern (speech/image) recognition, time-series prediction and modelling, function approximation, classification, adaptive control and other areas.

Most referred book on neural network defines it as follows [**16**]:*A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experimental knowledge and making it available for use. It resembles brain in two respects:*

- Knowledge is acquired by the network from its environmental through a learning process.
- Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.

3.3.2 Architecture of neuron model

A neuron is an information processing unit that is fundamental to the operation of a neural network. The block diagram of neuron model is shown in --, which forms basis for designing neural networks. It has three basic elements:

- A set of *synapses* or *connecting links*, each of which is characterized by a weight or strength of its own. Specifically, a signal x_j at the input of synapse j connected to neuron k is multiplied by synaptic weight w_{kj}. First subscript refers to the neuron in the question and second subscript refers to the input end of the synapse to which the weight refers.
- An *adder* for summing the input signals, weighted by the respective synapses of the neuron; the operations described here constitutes a *linear combiner*.
- An *activation function* for limiting the amplitude of the output of a neuron. The activation function is also referred to as a squashing function in that limits the permissible range of the output signal to some finite value. Typically the normalized amplitude range of the output of a neuron is written as the closed unit interval [0,1] or alternatively[-1,1].Basic activation function used are tan sigmoid, pure linear, hard limiter and log sigmoid. Based on application different activation functions are used.

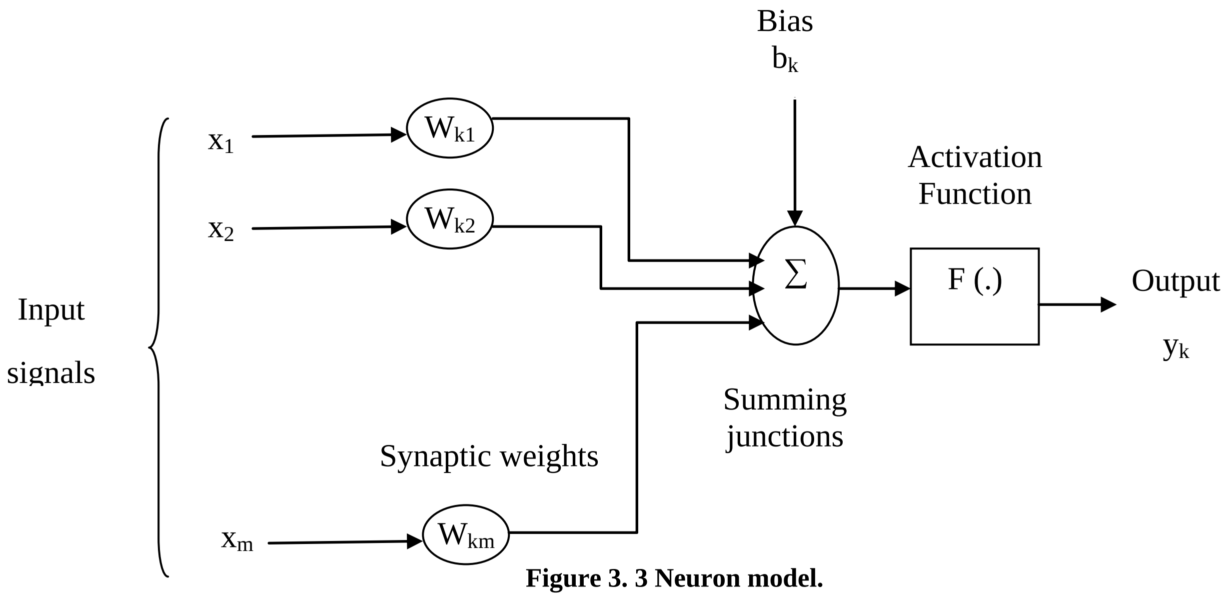

Figure 3. 3 Neuron model.

In mathematical terms, we may describe a neuron k by writing the following of equations:

$$y_k = F\left(= \sum_{j=1}^{m} w_{kj}\ x_j + b_k\right) \quad (3.1)$$

3.3.3 Neural Networks Architecture

The topology of a NN plays an important role for its achievable performance. Depending on the pattern of connections that a NN uses to propagate data among the neurons, it can be classified into one over two basic (non exhaustive) categories.

- *Feed forward networks* where data enters at the inputs and passes through the network, layer by layer, until it arrives at the outputs, with classical examples being the Perceptron [**18**] and Adaline [**18**].
- *Recurrent neural networks* that contain feedback connections, which are connections extending from outputs of neurons to inputs of neurons in the same layer or previous layers. In contrast with feed-forward networks, the recurrent network has a sense of history and this means that pattern presentation must be seen as it happens in time.

3.3.3.1 Feed forward networks architecture and operation:

As mentioned earlier FF networks includes single layer perceptron and Multilayer perceptron(MLP) and Adaline. Thesis describes MLP, since it used in Data rate prediction. Multilayer feed-forward networks(FF) consist of neuron units arranged in layers with only forward connections to units in subsequent layers. The connections have weights associated with them. Each signal traveling along the link is multiplied by the connection weight. The first layer is the input layer, and the input units distribute the inputs to units in subsequent layers. In the following layers, each unit sums its inputs and adds a bias or threshold term to the sum and nonlinearly transforms the sum to produce an output. This nonlinear transformation is nothing but the activation function of the unit. The output layer units often have linear activations. Activation function are discussed in previous neuron model The layers sandwiched between the input layer and output layer are called hidden layers, and units in hidden layers are called hidden neuron units. Such a 4 layer network is shown in fig.3.4. $x_i(n)$ represent the input to the network, f_j and f_k represent the output of the two hidden layers and $y_l(n)$ represents the output of the final layer of the neural network. The connecting weights between the input to the first hidden layer, first to second hidden layer and the second hidden layer to the output layers are represented by , w_{ij}, w_{jk} and w_{kl} respectively.

If P_1 is the number of neurons in the first hidden layer, each element of the output vector

of first hidden layer may be calculated as,

$$f_j = F_j\left[\sum_i^N w_{ij}\ x_i(n) + b_j\ \right] \quad , i=1,2,3\ldots N\ ,j=1,2,3\ldots P_1 \tag{3.2}$$

where b_j is the threshold to the neurons of the first hidden layer, N is the number of inputs and $F_j(.)$ is the nonlinear activation function of the neurons of the first hidden layer which is defined in Table.1. The time index n has been dropped to make the equations simpler. Let P_2 be the number of neurons in the second hidden layer. The output of this layer is represented as, f_k and may be written as

$$f_k = F_k\left[\sum_i^{P_1} w_{jk}\ f_j + b_k\ \right] \quad k=1,2,3\ldots P_{2}. \tag{3.3}$$

Where, b_k is the threshold to the neurons of the second hidden layer. The output of the final output layer can be calculated as

$$y_l(n) = F_l\left[\sum_i^{P_2} w_{kl}\ f_k + b_l\ \right] \quad l=1,2,3\ldots P_{3}. \tag{3.4}$$

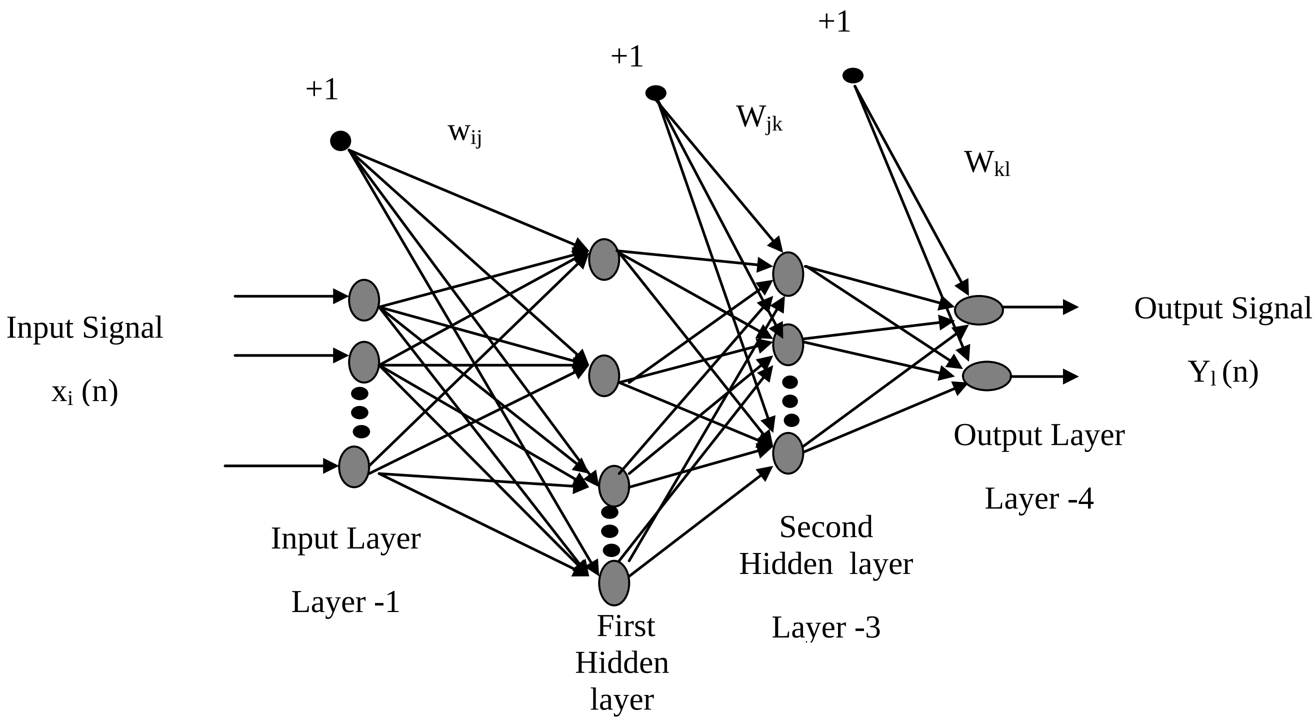

Figure 3. 4 MLP Architecture.

where, b_l is the threshold to the neuron of the final layer and P_3 is the number of neurons in the output layer. The output of the MLP may be expressed as

$$y_l(n) = \mathrm{Fn}\left[\sum_{k=1}^{P_2} w_{kl} F_k \left(\sum_{j=1}^{P_1} w_{jk} F_j \left\{\sum_{i=1}^{N} w_{ij}\, x_i\,(n) + b_j\right\} + b_k\right) + b_l\right] \quad (3.5)$$

Operation.

In any manifestation, a NN has to be configured such that the application of a set of inputs produces the desired set of outputs. This can be achieved by properly adjusting the weights w_{jk} of the existing connections among all (j, k) neuron pairs. This process is called learning or training. Learning can be generally distinguished between supervised and unsupervised learning (with reinforcement learning being also an option). In supervised learning, the NN is fed with teaching patterns and trained by letting it change its weights according to some learning rule, the so called back propagation rule [**18**]. The NN learns the input–output mapping by a stepwise change of the weights with the objective to minimize the difference between the actual and desired output. In the next step the actual output vector is compared with the desired output. Error values are assigned to each neuron in the output layer. The error values are back-propagated from the output layer to the hidden layers. The weights are changed so that there is a lower error for a new presentation of the same pattern. As a result of this procedure, the weights on the connections between neurons are properly adjusted so as to encode the actual knowledge of the NN. At that time, the NN can be used for the purpose that was initially set up for.

3.3.3.2 Recurrent Neural Networks (RNN)

As said previously these network differ itself from FF networks in that it has at least one feedback loop. They address the temporal relationship of inputs by maintaining internal states that have memory. RNNs have proven to be effective in learning time-dependent signals that have short term structure. For signals with long term dependencies, RNNs are less successful, since during training, the error gets "diluted" when passed back through the layers many times . Due to their dynamic nature, RNNs have found great use in time series prediction. In literature two types of recurrent networks can be found widely in use they are Elman and Hopfield network. Thesis discusses Elman network used for data rate prediction. Elman networks are two-layer back propagation networks, with the addition of a feedback connection from the output of the hidden layer to its input. This feedback path allows Elman networks to learn to recognize and generate temporal patterns, as well as spatial patterns. And it also detects and generates time-varying patterns. Elman network is shown in figure 3.5.

In Elman's recurrent network feedback connections layer is called context layer. These context layer store the outputs of hidden neurons for one time step and feed them back to the input layer.The inputs to the hidden layers are combination of the present inputs and the outputs of the hidden layer which are stored from previous time step in context layer. Hence the outputs of the Elman network are functions of present state, previous state (that is stored in context units) and present inputs. [**19**] Let each layer has its own index variable, *k* for output nodes, *j* (and *h* for recurrent connections) for hidden nodes and *i* for input nodes. The input vector is propagated through a weight layer **V** and combined with the previous state activation through an additional recurrent weight layer, **U**. The output of *j* th hidden node is given by

$$v_j = F(a_j(n)) \tag{3.6}$$

$$a_j(n) = \sum_i x_i\,(n)\,v_{ji} + \sum_h v_h\,(n-1)u_{jh} + b_j \tag{3.7}$$

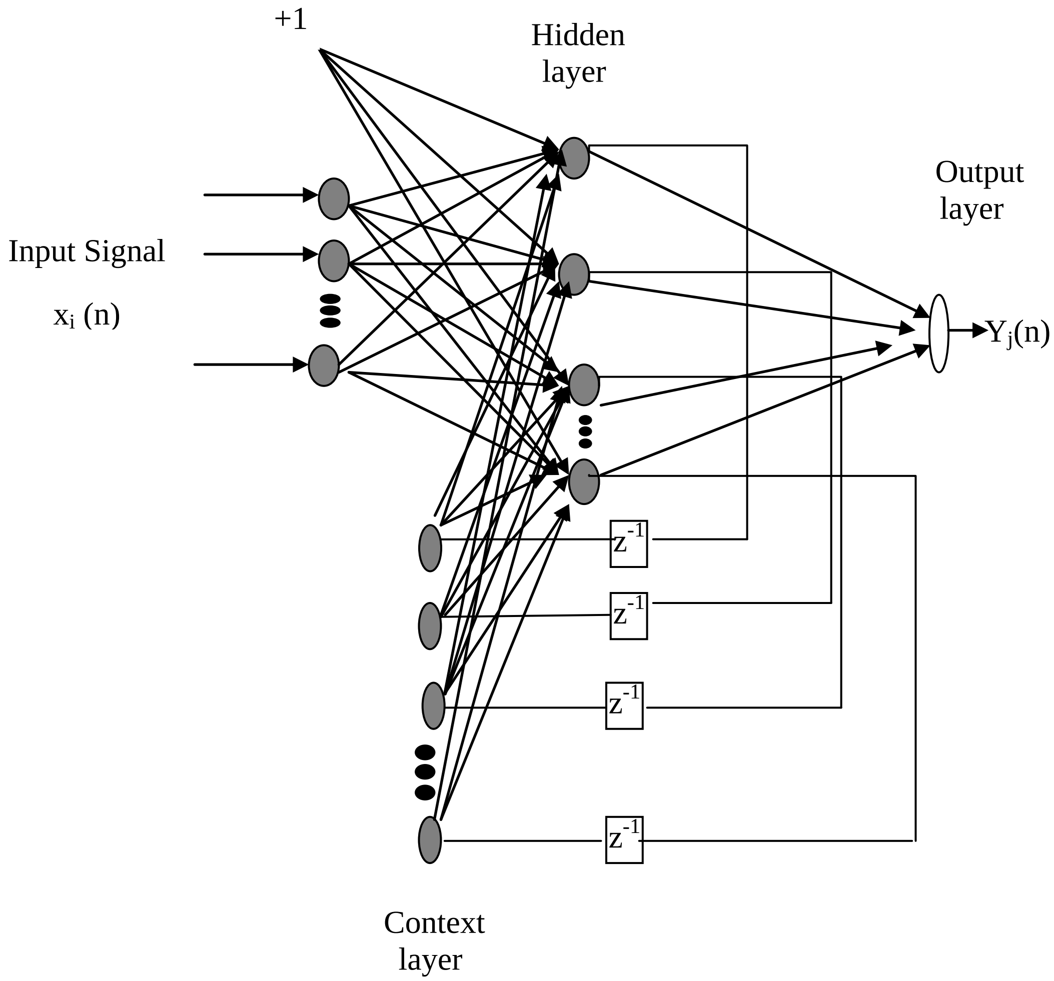

Figure 3. 5 Elman neural networks.

recurrent connections) for hidden nodes and *i* for input nodes. The input vector is propagated through a weight layer **V** and combined with the previous state activation through an additional recurrent weight layer, **U**. The output of *j* th hidden node is given by

$$v_j = F(a_j(n)) \tag{3.6}$$

$$a_j(n) = \sum_i x_i\ (n)\, v_{ji} + \sum_h v_h\ (n-1) u_{jh} + b_j \tag{3.7}$$

and a_j is output of *j*th hidden node before activation. x_i is the input value at *i* th node. *bj* is the bias for *j*th hidden node, and *F* is the activation function. Tan sigmoid activation function is hidden nodes. The output of the Elman's network is determined by a set of output weights, **V**, and is computed as,

$$y_k(n) = \mathrm{F}(\mathrm{a_k(n)}) \tag{3.8}$$

Where $y_k(n)$ is the final estimated output of kth output node. Main advantage of Elman network is, it can be trained using backpropogation algorithm similar to feed forward network. It is dynamic network which mainly used in problems like time series predication.

3.4 Motivation for Data rate prediction using Neural Network

Here investigation is done on neural network which was proposed by [**4**] are suitable for accommodating in CR.As said before NNs can be used in any phases of cognition cycle. Thesis focus is on the channel estimation and predictive modeling phases and we present a benchmarking work that aims at evaluating the applicability of multiple types of NNs in the learning module of the cognitive engine within a cognitive terminal (see Fig3.2). The discussed

NN-based learning schemes should relax the reasoning process and assist in the optimum decision regarding the radio-configuration settings (mainly PHY and MAC layer) that provide the best QoS for the given problem and user/ application needs. It should be noted that QoS optimization is a multi-objective problem that depends on many quality metrics with dependent relationships, including bit error rate, frame error rate, power consumption, latency, data rate, etc., and as such, it should call for Pareto optimality, which balances the trade-offs among the multiple objectives. Nevertheless, the focus here is given only on one objective: the data rate. This is also aligned with objective, namely to showcase the feasibility of the discussed learning schemes, parameters. But these neural network schemes are used in next chapter to compare with ANFIS based learning scheme which is thesis contribution. As proposed in [**4**] NN can be used

- To learn from information measured by the terminal during the radio scene analysis
- To provide in the output the data rate that is most anticipated to be obtained per radio configuration (RAT/frequency), thus behaving as a predictor of the next expected data rate.

What is gained is that by associating each configuration with a predictable, achievable data rate, NN-based learning schemes may facilitate the cognitive terminal in making its decision regarding the configuration in which it should operate, selecting the best among a set of candidate ones. Accordingly, two neural network-based learning schemes have been set up and tested: the 'basic' and the 'extended' one. In both cases, multiple types of NNs with a considerable number of adjustable parameters have been investigated. These trial and error processes that were conducted in order to derive the best possible network in both basic and extended cases, are described and discussed more analytically in the following sections.

3.5 Basic NN-based Data Rate Prediction [4]:

3.5.1 Preparation procedure:

In the work proposed in this thesis, NN-based scheme is tuned in an arbitrary radio configuration, e.g. IEEE WLAN 802.11g. In order to exhibit the applicability of such NN-based learning schemes, an algorithm that will be used to train variously parameterized NNs so as to predict the data rate to be obtained by the configuration being under investigation. This algorithm aims at defining a target data rate for each of the input value(s) presented in the NN [**4**].

- *Let* $R = \{r_k\}$, *be the time-series collected by the radio-scene analysis (environment sensing) phase, where each element* r_k *represents a data rate value at time slot k,* $K \in N$.
- r_k *values are quantized in predefined reference values from a finite set* $M = \{m_1, m_2 . . . m_{|M|}\}$.
- *A time window of n slots is used to represent past experience and knowledge collected by the NN and is depicted in Figure.3.6.*
- *At any time k, the NN is fed with an input sequence* $R^{in} \subseteq R$ *the length of which equals the size of the time window, i.e.,* $R^{in} = \{r_i\}$, $i = 1, \ldots, n$.

- *In addition, in each slot within the time window the corresponding value r_i is associated with a weight, β_i, $i = 1. . . n$. (not be confused with the weights of the connections among neurons)*

- *Requiring that recently collected values should have greater weights, an exponentially weighted moving average with a smoothing factor a is used to configure β_i,, i.e.,*

$$\beta_{i=} a.(1-a)^i \qquad (3.9)$$

- *The objective of the above algorithm is to exponentially decrease the weighting for each older data value, giving much more importance to recent observations, while still not entirely discarding older observations.*

- For each r_k there will be a target data rate value, r_k^{tgt} , that will be used to train the basic NN at each time k and is derived as follows. Consider the above specific window of n slots.

- *Each of the reference values m_j, $j = 1...|M|$ is associated with a weight β_{mj} given by*

$$\beta_{mj} = \{ \sum {}^{n}{}_{i=1} \beta_i \ / \ r_i = m_j \ \} \qquad (3.10)$$

- *Weights β_{mj} actually represent the number of occurrences of each of the reference values m_j in M within the time window.*

- *The target data rate value corresponding to the input time sequence within the time window will be given by following relations.*

$$r_k^{tgt} = \arg_{mj} \max \ \beta_{mj} \qquad (3.1\ 1)$$

In other word, the target value selected is the one that has the maximum weighted sum, within the time window.

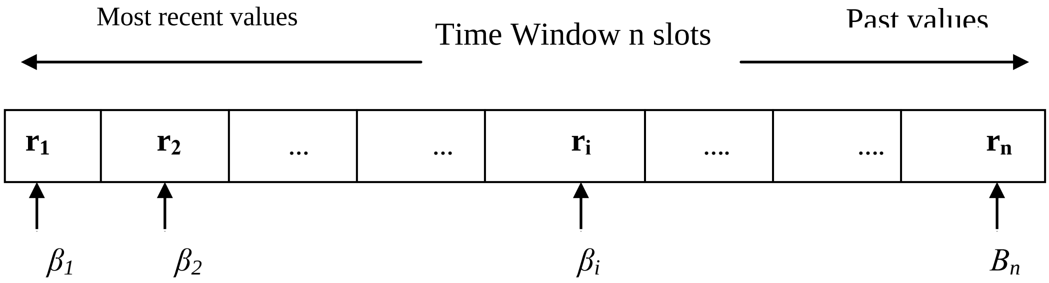

Figure 3. 6 Time window

3.5.2 NN Pattern selection for Basic Scheme

Here focus is given on the selection of the NN .Selection is based on network that gives best performance in terms of minimizing a predefined metric such as mean square error MSE and root mean square error (RMSE) and also percentage of prediction accuracy. Neural network that has been selected for basic scheme is Elman recurrent network. As discussed earlier it is two-layer back-propogation, recurrent network, with addition of a feedback connection from the output of the unique hidden layer to the input layer. As discussed earlier recurrent connection allows Elman network to both detect and generate time varying patterns. Elman networks are effective in learning time dependent signals that have short time structure. For signals with long term dependencies, ENNs are less successful, since during training, the error gets "diluted" when passed back through the layers many times.

Since in basic scheme it is considered whole day data rate prediction. In this case, time window length taken is small. Based on this other Feed forward networks are not tested since Elman network perform better in small structure time series prediction. Work on Elman network is previously done in reference [**4**]. So we set all parameters neural network as in [**4**]. The NN uses the "*tansig* " activation function, for the neurons in its hidden (recurrent) layer, and " *logsig"* function is used for the neuron in its output layer, respectively. A delay line of five slots has been inserted in the input layer, which corresponds to the time window. As mentioned in previous work Elman NN with 15 hidden nodes performed better in terms MSE. So its architecture is shown in Figure .3.7.

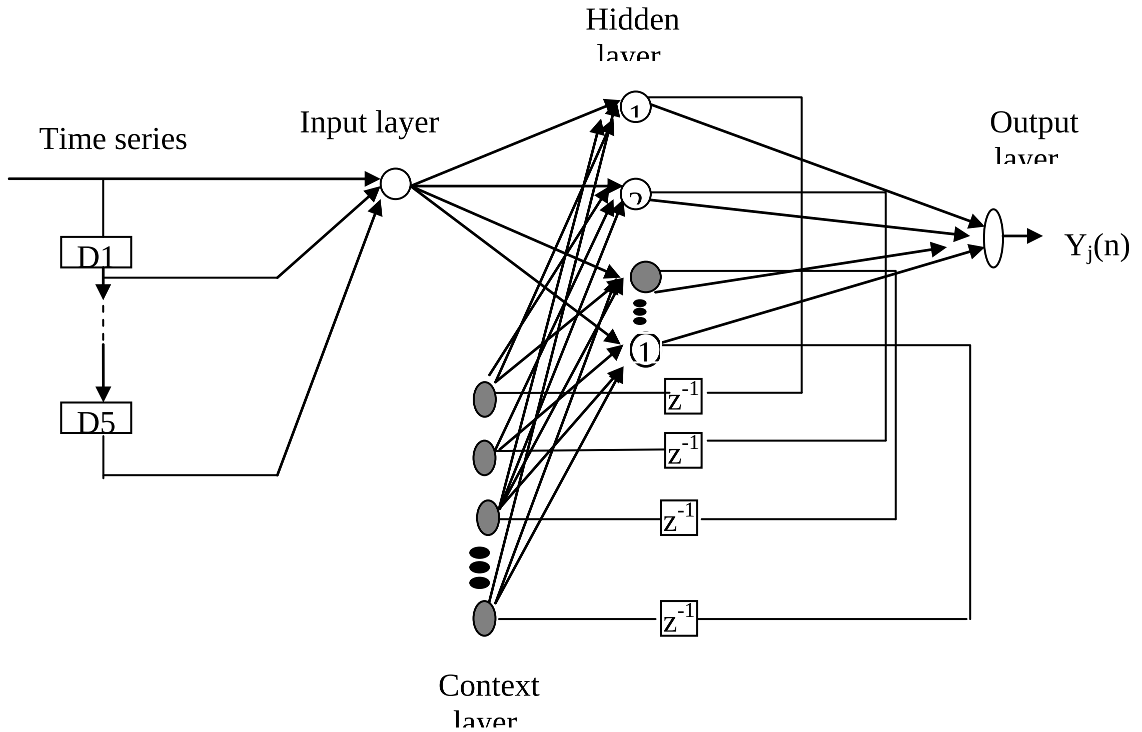

Figure 3. 7 Neural network for the basic learning scheme.

3.5.3 Simulation Results and Discussion.

For simulation M values has been set to 6 reference bit rate values(in Mbps) $|M| = 6$ as follows: m_1=6, m_2=12, m_3=24,m_4=36,m_5=48,m_6=54i.e., the values that might correspond to the data rate obtained by a typical WLAN-equipped terminal. The time window equals to n = 5. The smoothing factor a of the exponential moving average algorithm is arbitrarily set to a = 0.362, thus resulting to the respective calculated weights $\{\beta_i\}$= {.1488, .1217, 0.0996 ,0.0814, 0.0666}.The time-series R includes values from the M set, which are randomly generated according to a selected distribution function, depicted in Figure 3.8 (normal line), that assigns bigger probability to the appearance of m_1 = 6. The target values r_k^{tgt} are calculated according to Eq. (3.12) .

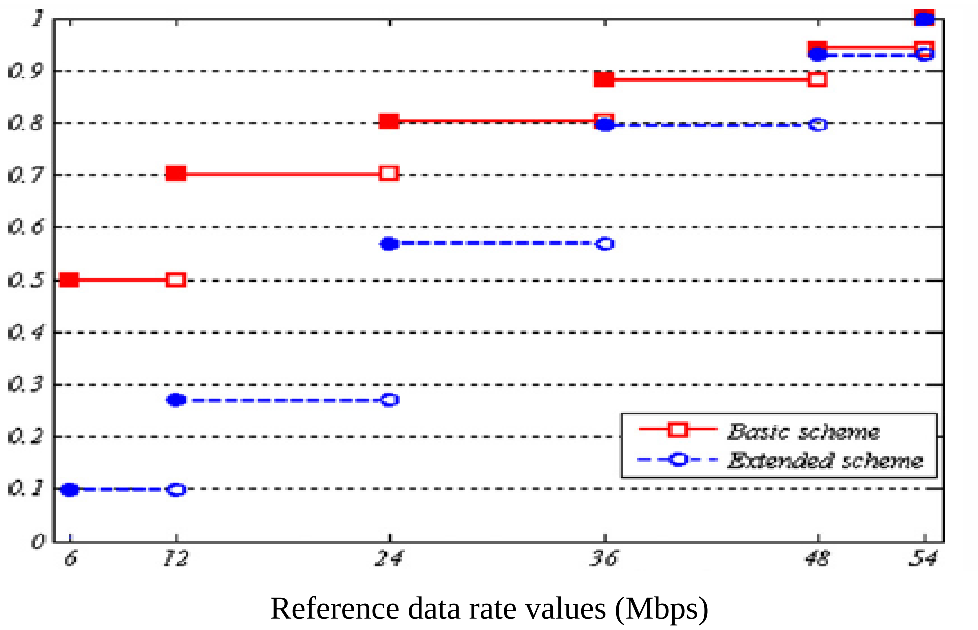

Figure 3. 8 Cumulative distribution functions of input time-series [4**].**

Thesis considers training parameters as of reference [**4**]. Elman network is tested with different number of hidden layers by keeping training data size set and training parameters same as previous reference's best case. For the training session, the input and target values have been properly normalized in the range of [0, 1] in a pre-processing phase . During training, weights and bias values have been updated according to a gradient descent momentum and an adaptive learning method (traingdx in [**18**]).As stated before MSE and accuracy of Data rate prediction percentage have been used as a metric for measuring the performance of Elman network.

For analysis two data sets are used, which are extracted from the whole input sequence and serve as target values for teaching the NN:

- *A "training set" (seen data) which is used to build the model i.e. determine its parameters, during the so called training session.*
- *A "Validation set" (unseen data) which is used to measure the performance of the network by holding its parameters constant. Term "unseen" refers the data that have never been used to update the weights of the network.*

The importance of testing the network with both datasets, when searching for the best structure, is significant, since a small error in the training set can be misleading. If the network has not been trained well, it may not learn the basic structure of the data, but rather learn irrelevant details of the individual cases, overfitting the training data or *overtraining*. This would lead to a small error during testing with the training set, but in a large error during testing with the validation dataset. In general, performance on the training only tells us that the model learns what it's supposed to learn, but it is not a good indicator of performance on unseen data, i.e. whether the NN is able to generalize well or not [**16**].

Moreover, the number of hidden layers and/or neurons plays a critical role in the learning process and strongly influences the performance of the network. The use of too few hidden neurons would result in a NN that is unable to learn what we want it to learn. On the other hand the use of too many hidden neurons would dramatically increase the time needed to learn, without yielding any significant improvement in the performance of the network and which leads to overfitting. There exist some valid rules to set the number of hidden nodes [**18**]but in general, it is better to start with a big net, train, and then carefully follow a pruning strategy for gradually reducing the size of the network.

After completion training, the performance of NN has been tested in both the "known" and "unknown" sequences comprising 100 data points each. The known sequence is actually a subset of the "training" set. Also, in order to measure NN's degree of generalization [**16**]. A completely unknown sequence in order to constitute the so called validation set or validation sequence. During the validation, the MSE between the value produced by the NN and the expected target value has been recorded. An acceptable NN design pattern should satisfy following criteria:

- *$(MSE_{trn} \leq MSE_{thres}) \wedge (MSE_{val} \leq MSE_{thres})$, where MSE_{trn} is the final MSE produced during the training session, MSE_{val} – is the final MSE produced during the validation and MSE_{thres} –is a desirable upper threshold for MSE which is arbitrarily set here to 0.02*
- *Minimize* $|MSE_{trn} - MSE_{val}|$.

The first criterion is self-explanatory. Regarding the selection of the MSE threshold, the network has been requested to run for a number of epochs sufficient to lower the MSE to a little amount (MSE goal). In following simulation analysis MSE_{thres} is set 2% or 0.02. So MSE value higher than threshold results in lower performance in terms of how well the NN

“ learnt its lesson”. As for the second criterion, it is used here in order to guarantee a certain level of generalization, meaning that the neural network must have the ability to behave efficiently when dealing with unseen input data and thus avoiding over fitting the training data.

Here first presents simulation results performance for different number of hidden layers and decides best among them to be used for Data rate prediction. All training case and validation case results are tabulated in Table 3.1. It includes Number hidden layer taken, MSE in training and MSE in validation and percentage of data rate prediction accuracy. Also figure 3.9, 3.10 and 3.10 depicts MSE curve, data matching between target values and predicted values in training case and validation case for best performing NN respectively. All simulations are run for 500 epochs and learning rate is set to 0.001. As previously mentioned *tansig* activation function is used in its hidden layer and *logsig* activation function in the output layer.

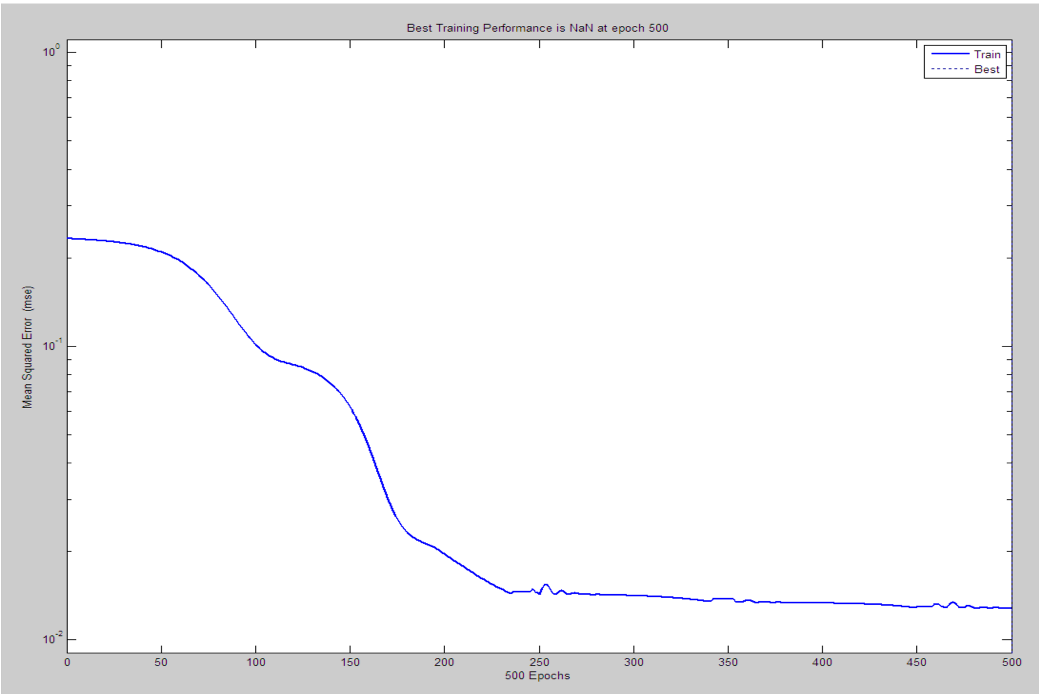

Figure 3. 9 MSE curve for -basic scheme.

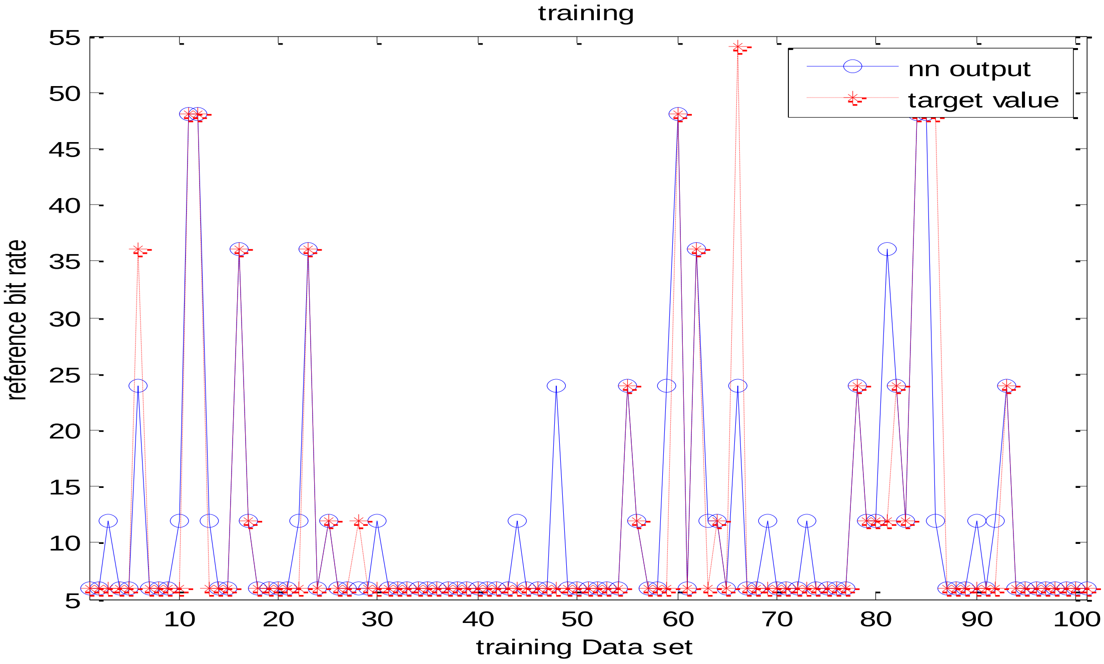

Figure 3. 10 Prediction accuracy of selected NN in training sequence-basic scheme.

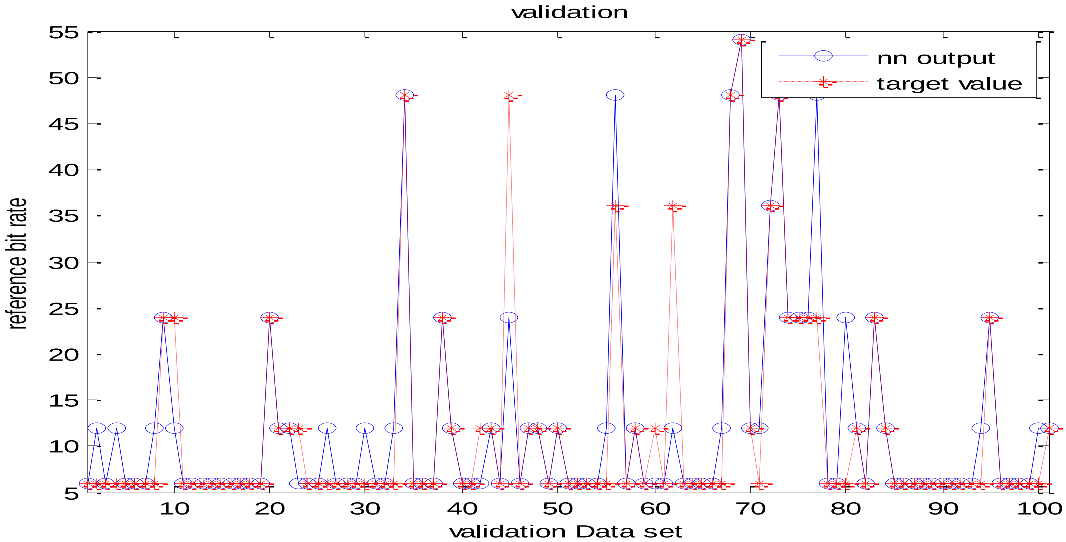

Figure 3. 11Prediction accuracy of selected NN in validation sequence-basic scheme.

From above table it can be concluded that, NN having 15 hidden node performed better in terms prediction accuracy and MSE difference. Same network is depicted in figure.3.8. As it can be observed, the MSE produced during the validation naturally exceeds slightly the one produced during the training. When fed with the known sequence, the NN actual output seems to follow the target values (that are expected according to the input that feeds the NN), giving very few errors, which shows that the network has been trained well. The same applies for the unknown sequence. The NN performs well during the validation session and it can be observed that the network has learned the basic structure of the data but at the same time it is also able to generalize well.

Table 3. 1 Performance index for different NN.-basic scheme

No hidden nodes	MSE_{train}	MSE_{valid}	$\lvert MSE_{train}\ MSE_{valid} \rvert$	RMSE	RMSE	Prediction accuracy in training case	Prediction accuracy in validation case
2	0.0212	0.0161	0.0051	0.1260	0.1456	62	52
5	0.0165	0.0202	0.0037	0.1456	0.1422	73	71
10	0.0105	0.0130	0.0025	0.0795	.1140	81	75
15	0.0104	0.0 128	0.0024	0.1161	0.1020	83	81
20	0.0131	0.0128	0.0030	0.1125	0.1134	76	71

3.6 Extended NN -based Data Rate Prediction [4]:

3.6.1 Preparation procedure:

For the extended NN scheme, the complexity of the problem is raised by further taking into account and co-estimating a ‘‘time zone” parameter. It is assumed that the day is divided into time zones and that during each of them; the configuration in question is associated with a mean, most usually observed data rate value, which is denoted as $\bar{m}_{tz} \in M$. This value is enhancing NN learning scheme with a feature of past experience. Let $R^{ext} = \{r_k^{ext}\}$, K∈N be the new time-series collected by the radio-scene analysis (environment sensing) phase. As in the basic scheme, a time window of n slots is considered and at any time k, the NN is fed with an input sequence $R^{in,ext} \subseteq R^{ext}$ the length of which equals the size of this time window, i.e., $R^{in} = \{r_i^{ext}\}$, i= 1…n. For each pair comprising an input time sequence $R^{in,ext}$ of length n and a specific time zone of the day, within which the NN is expected to operate, there corresponds a new target data rate value $r_k^{tgt,ext}$. This value will be used for supervising the training and is calculated at each time k as follows:

Step 1: Temporary target value $r_k^{tgt.tmp}$ is calculated as in Eq. (3.12) by applying $R^{in,ext}$ as input.

Step 2: The distance (absolute difference), dst_k, between the target value at k i.e. $r_k^{tgt,tmp}$ and mean value $\overline{m}_{tz}$, which corresponds to the considered time zone is calculated. Specifically and assuming that M is the set defined in Sub-section 3.6.3, if the target value . $r_k^{tgt,tmp}$ and mean value $\overline{m}_{tz}$ equals $m_i \in M$, $1 \leq i \leq |M|$ and the mean value as $\overline{m}_{tz}$ equals $m_j \in M$, $1 \leq j \leq |M|$, then the distance dst_k is taken equal to $|i - j|$, $1 \leq i, j \leq |M|$, For example, if $r_k^{tgt,ext}$ equals m_1=6 Mbps and mean value $\overline{m}_{tz}$ equals $m_{2=}$12 Mbps ,the distance dst_k equals $|1 - 2|$=1; similarly the distance from m_1 = 6 Mbps to m_3 = 24 Mbps equals equals $|1 - 3|$=2 and so forth.

Step 3: The new weights $\beta_i^{'}$ (where i = 1. . . n and n is the number of slots in the time window) are recalculated with the use of an exponentially weighted moving average function, with a smoothing factor a_k=$1 - \chi^{dst_k}$, $\chi \in (0..1)$. For example, assuming that M is the set defined in Sub-section 3.6.3 and that n = 8, figure 3.12 illustrates the weight values per time slot when dst_k = 1. . . 5. The above rule is used so that greater distance between the calculated value and the target value should lead to a lower slope of weight decrease. A lower slope, as shown in figure 3.12, gives high importance (high weights) to past observations so as to eliminate the abovementioned distance. The slope of weight decrease is expressed by the smoothing factor a.

step 4:Finally $r_k^{tgt,tmp}$ is calculated from Eq.(3.12) by using the new weights $\beta_i^{'}$ and $R^{in,ext}$

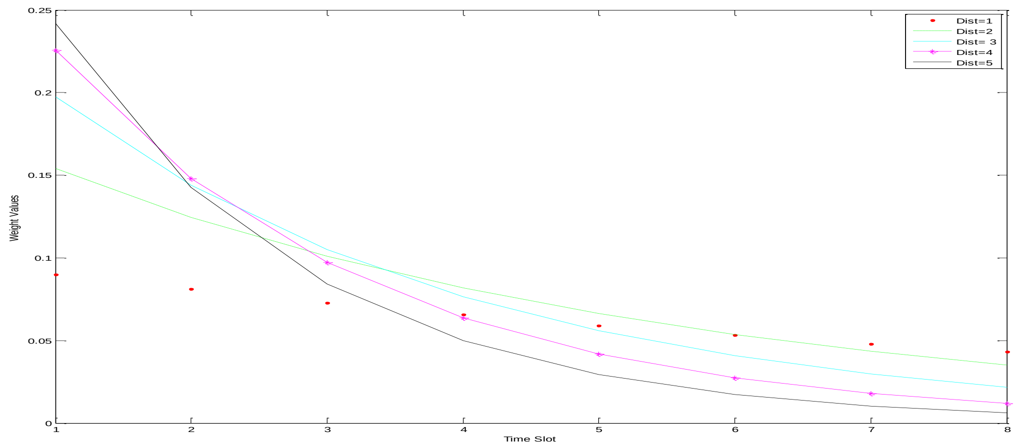

Figure 3. 12 Weight values per time slot for the extended scheme.

3.6.2 Results and discussion

In extended case as said previously complexity problem is raised by considering time zone parameter. Here also for selection of NN design pattern, the focus is as before, on a specific, arbitrary radio configuration, e.g. WLAN 802.11g. It is assumed that the number of reference bit rate values is $|M|$=6. Including Elman NN different NN pattern are tested, because complexity problem is increased. Here feed-forward back propagation (FF) NNs with different hidden layers and focused time-delay neural network (FTDNN) are considered for analysis. FF NNs are discussed previously. FTDNN is a feed-forward input- delay back-propagation network, which consists of a feed-forward network with a tapped delay line at the input. FTDNN is a network well suited to time-series prediction. For simulation analysis time window has been set n=8 and $\chi = .7$.

It is also assumed that the day is divided in four equal time zones as follows: 06:00–12:00, 12:00–18:00, 18:00–24:00 and 00:00–06:00. In each of these time zones a different mean value $\overline{m}_{tz}$ is observed, let them be set equal to 24, 6, 36 and 48 Mbps for each of the four time zones, respectively; this might reveal for instance the existence of high load situation during the mid working day. R^{ext} includes values from the M set which are randomly generated according to a selected probability distribution function, depicted in figure 3.8 (dotted line), that assigns bigger probability to the appearance of $\overline{m}_{tz}$ depending on the time zone. The target values $r_k^{tgt,tmp}$ are calculated by following the steps mentioned in the previous sub-section. The NN uses the *tansig* function of table.1 for the neurons in its hidden (recurrent) layer, and the *purelin* for the neuron in its output layer, respectively. For the training session, the input and target values have been properly normalized in the range of [-1, 1].A number of different cases has been tested to evaluate the extended NN scheme. Table 3 gives an overview of the parameters used to define those test cases and performance index in terms of MSE, RMSE and prediction accuracy.

For the first set (test cases 1–4), as presented in Table.3.2 a feed-forward back-propagation (FF) NN has been used. Here test cases are according time zone as previously mentioned .First case 06:00–12:00 with mean data rate set equal to 24 Mbps ,second case 12:00–18:00 with mean value 6 Mbps, third case 18:00–24:00 with mean set equal to 36 Mbps and fourth case 00:00–06:00 with mean set equal 48 Mbps respectively. All the neural networks are tested for four cases. In the first two cases the network consisted of 10 hidden nodes in the hidden layer, while in the second two cases it consisted of 15 hidden nodes. The networks have been trained with function has been used for updating the weights and biases. For the training session, 1000 sample taken for each of the time zone.

Next case FTDNN network is considered for testing. This network is feed-forward input- delay back-propagation network, which consists of a feed-forward network with a tapped delay line at the input. As said previously network is well suited to time-series prediction. The delay line has been set to 8. As presented in table .3, for first time zone, third time zone and fourth time zone respectively best results are found with 10 hidden layers, whereas for the second zone requires 15 hidden neurons .The Levenberg-Marquardt optimization function has been used for the training. Also, the same 1000 sample data points as in the previous test set have been used for training.

At last, table presents results of Feed-forward back-propagation with multilayer. Here two layer and three layer FF networks are considered. The networks have been configured to have two hidden layers, in case 13to 15. From 16-20, they have been configured to have three hidden layers. In general, using more than one hidden layer is almost never beneficial. The only situation in which a NN with two hidden layers may be required in practice is when the network has to learn a function having discontinuities. For two hidden layers case hidden neurons are set 10 neurons each , while in the three layer case hidden neurons are set to 8-10-10 neurons, respectively. All networks have been configured to have a tapped delay line of 8 slots. The networks have been trained with the use of Bayesian regularization back-propagation, which is believed to produce networks that generalize well.

The training lasted for 500 epochs and the learning rate of 0.0001 has been used. The input has been the same 1000 data samples, for each time zone. The best available network design pattern is the one corresponding to the 16th test case in Table 3, since it is the one that produces best percentage of prediction accuracy and satisifies optimum MSE criteria similar to those in Sub-section3.5.3. This case designates a feed-forward back-propagation network with two hidden layers with ten *tansig* nodes each, and a *purelin* node in the output layer (Figure.3.13). The training session has lasted for 500 epochs and a learning rate of 0.0001 has been used. Finally, a set of 1000 training data input values have been used with a tapped delay line of 8 slots. In the sequel, the trained extended-NN has been tested in both a known (subset of training set) and an unknown (validation) sequences comprising 100 data points each. In the case of the known sequence, the NN produces an MSE = 0.0022, while in the case of the validation sequence, the MSE = 0.0184. Figures 3.14 and 3.15 illustrate the prediction accuracy in case of training and validation case. Again, as in the case of the basic scheme, the MSE produced during the validation naturally exceeds slightly the one produced during the training. The output of the NN during both cases is very close to the target values, which produces a very small error. Due to the complexity of the problem (multiple time zones), a two hidden-layer network performs better.

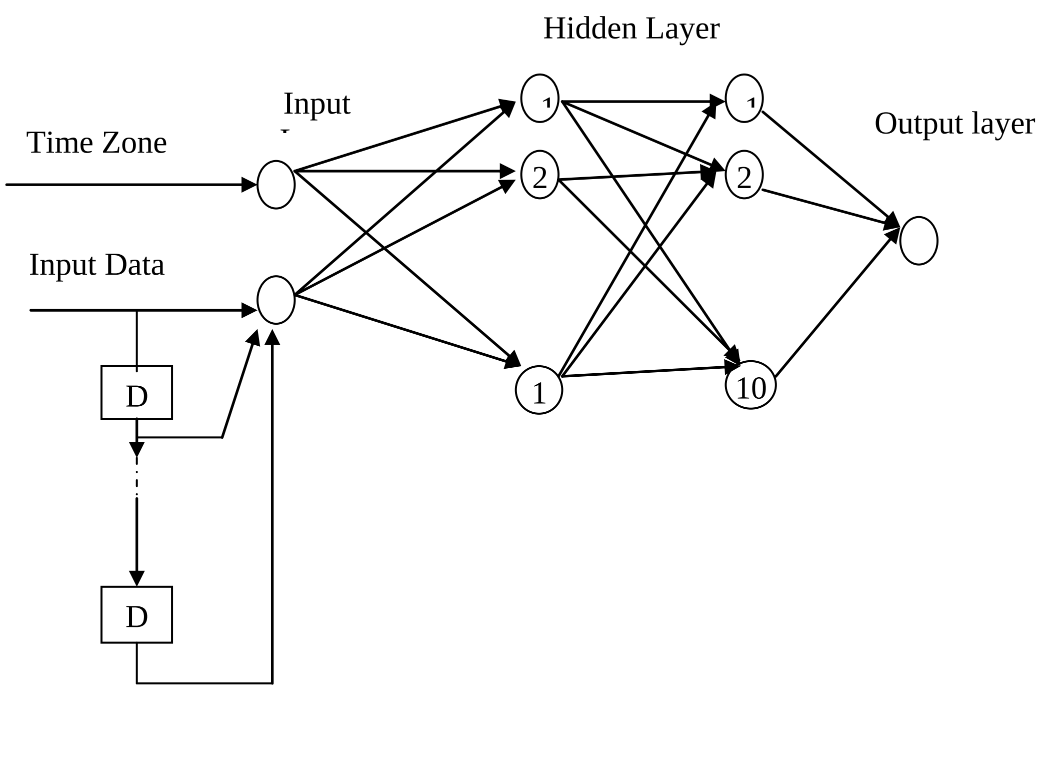

Figure 3. 13 Neural network for extended scheme.

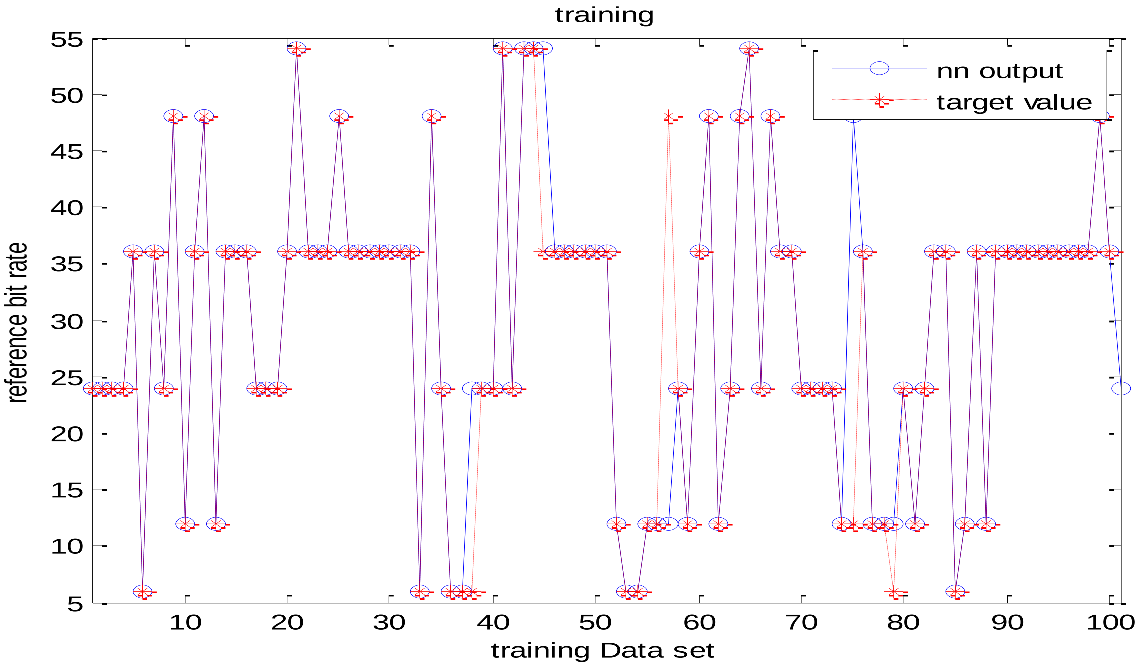

Figure 3. 14 Prediction accuracy of selected NN in training sequence-Extended scheme.

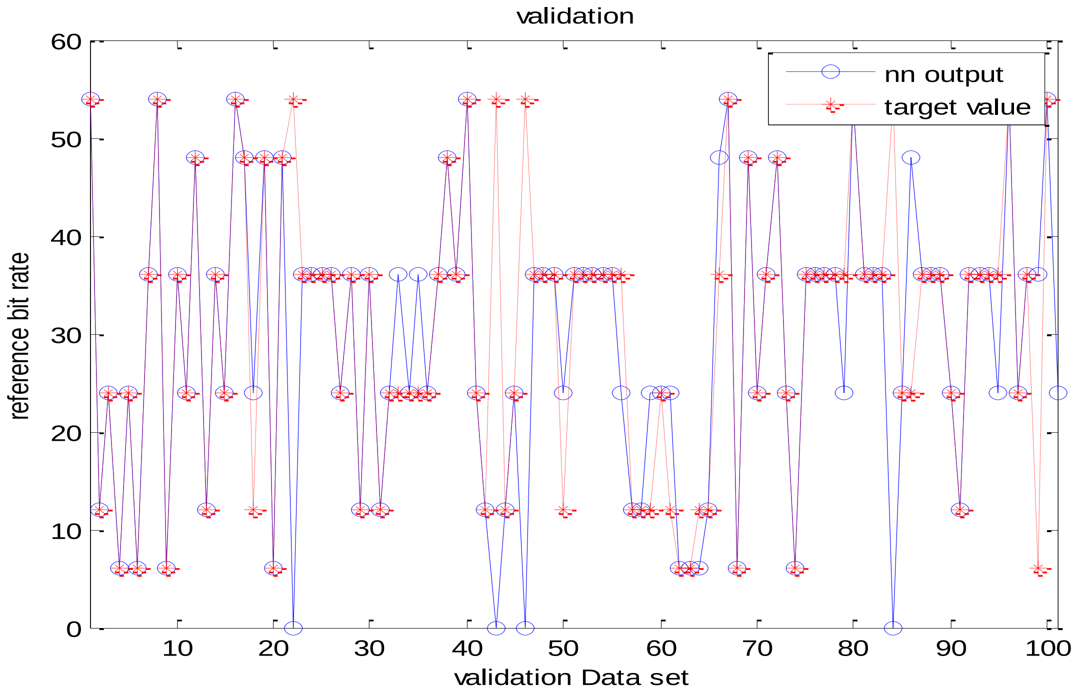

Figure 3. 15 Prediction accuracy of selected NN in validation sequence-Extended scheme.

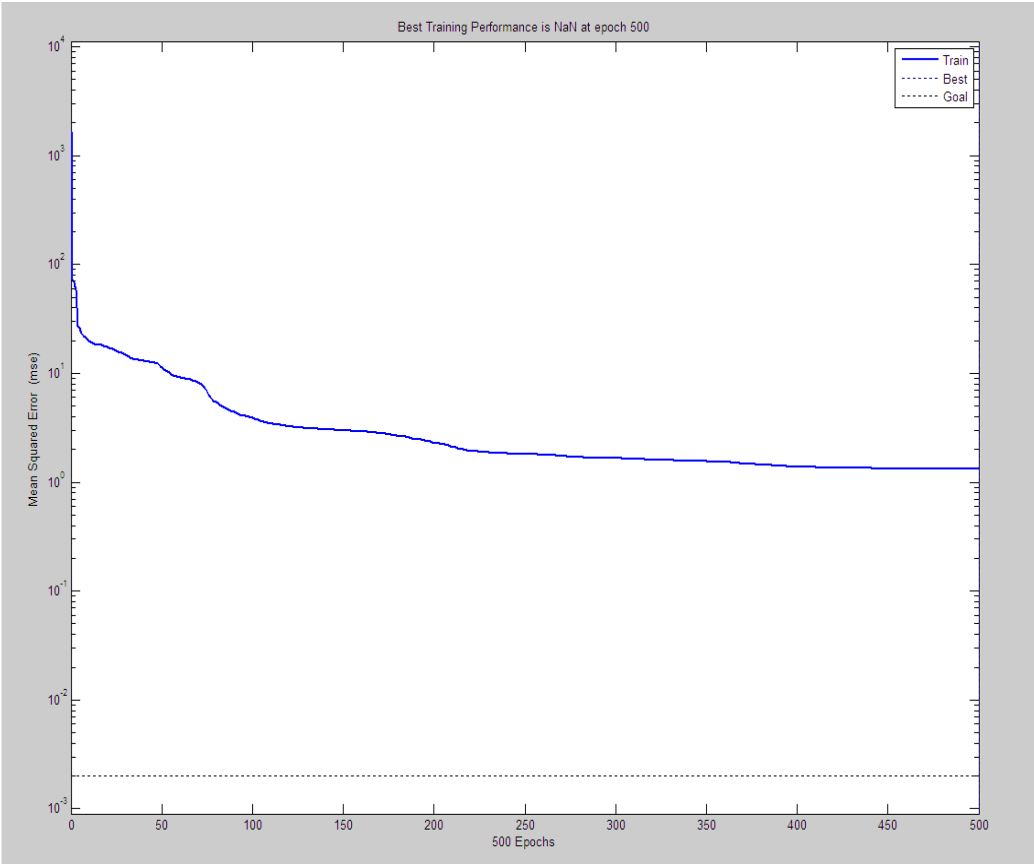

Figure 3.16. MSE curve for –extended scheme

This observation can be generalized for all cases. The performance of the NN is dramatically increased when the number of hidden layers is increased.This seems logical, since smaller networks don't have the ability to distinguish between the time zones (separate the problem). Conversely, adding more neurons into the two hidden layer network does not raise the performance of the network. Actually, the error increases when more hidden neurons are used and also prediction accuracy is also less. This is normal since there is a theoretically best performance that cannot be exceeded by adding more neurons; the network learns irrelevant details of the individual cases. So NNs scheme proposed by K. Tsagkaris , A. Katidiotis, P. Demestichas [**4**] can generalize well, and giving output values very close to the target values with less error.

Table3. 2 Performance index of NNs - extended case

S.N	Types of NN	Hidden neurons	Hidden layers	MSE_{trn}	MSE_{valid}	$MSE_{train} - MSE_{valid}$	$RMSE_{train}$	$RMSE_{valid}$	Prediction Accuracy training	Prediction accuracy validation
1	FF	15	1	0.0051	0.0271	0.022	0.1115	0.1646	84	77
1	FF	15	1	0.0104	0.0186	0.0082	0.1186	0.1362	71	60
3	FF	10	1	0.0068	0.0242	0.0174	0.0884	0.1554	8 3	63
4	FF	10	1	0.0098	.0240	0.0142	0.1130	.1549	87	73
5	Elman	15	1	.0175	0.0170	0.0005	0.1082	0.1304	71	71
6	Elman	15	1	0.0158	.0102	0.0056	0.1225	.1011	76	73
7	Elman	10	1	.0164	.0156	0.0008	0.1233	.1248	73	70
8	Elman	15	1	.0214	.0181	0.0033	0.1582	.1347	70	66
9	FTDNN	10	1	0.0064	0.0079	0.0015	0.0967	0.0889	83	79
10	FTDNN	15	1	0.0071	.1098	0.1027	0.115	0.3314	83	75
11	FTDNN	10	1	0.0078	.0107	0.0029	.00951	.1036	84	74
12	FTDNN	10	1	0.0091	.3537	0.3446	0.0637	0.5947	85	78
13	FF	10-10	2	0.0015	0.0293	0.0278	0.0936	0.1713	90	86
14	FF	10-10	2	0.0017	0.1860	0.1843	0.1089	.0.4312	90	84
15	FF	10-10	2	0.0017	0.0228	0.0211	0.1335	0.1511	89	83
16	FF	10-10	2	0.0022	0.0210	0.0186	0.0903	0.1516	92	88
17	FF	8-10-10	3	4.3233e-004	0.0836	0.0832	0.1235	0.2891	88	81
18	FF	8-10-10	3	0.0011	0.0214	0.0203	0.0642	0.1462	83	82
19	FF	8-10-10	3	5.0815e-004	0.0189	0.0184	0.1061	0.0941	86	79
20	FF	8-10-10	3	2.7478e-004	0.0179	0.0176	0.0859	0.1338	85	80

3.7 Conclusion

This part of work is previously proposed in [**4**]. Same work is simulated with slight parameter modification and found that thesis results are almost matched to previous work of Tsagkaris. So NNs can be used for assisting Cognitive radio. In this case multi objective problem was made single objective and analyzed. As it seen in extended case when the complexity of problem increased it led use of complex multilayer structure. As the layer number increase mathematical complexity also increase and learning phase becomes slow. CR has to deal with multi objective problems so complex NN may not be good solution. So in next chapter thesis proposes ANFIS based assistances to cognitive radio which uses rule based expert system with neural network training ability

CHAPTER 4

ANFIS BASED DATA RATE PREDICTION

4.1 Introduction

The previous chapter of the thesis presented NNs based data rate prediction for cognitive radio. In this chapter proposes ANFIS based data rate prediction for cognitive radio is proposed. ANFIS is a hybrid AI technique, which combines best features of Fuzzy logic and parallel processing neural networks. It possesses fast convergence and has more power than BP neural network. Various form of ANFIS methods are explored for data rate prediction. ANFIS methods are comparatively good at prediction than Complex neural networks.

Chapter covers basic overview of ANFIS, Fuzzy-c means clustering based ANFIS ,subtractive clustering based ANFIS and these application to “basic” and” extended” based data prediction. At last comparative study between NN method and ANFIS methods are done. It is shown that ANFIS method is good candidate to assist cognitive radio.

4.2 (ANFIS): An overview

4.2.1 Basics of Fuzzy Modeling

Earlier system modeling problems were solved using conventional mathematical tools (like differential equations), but they are not well suited with vaguely defined and uncertain system. Takagi and Sugeno [**20**] introduced fuzzy inference system which employs fuzzy if – then rules that can model the qualitative aspects of human knowledge and reasoning processes without employing quantitative analyses. So Fuzzy modeling, has found numerous practical applications in control, prediction and inference. Fuzzy logic is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is approximate rather than precise. In contrast to binary sets having binary crisp logic, the fuzzy logic variables may have a membership value of not only 0 or 1. Just as in fuzzy set theory with fuzzy logic the set membership values can range (inclusively) between 0 and 1, in fuzzy logic the degree of truth of a statement can range between 0 and 1 and is not constrained to the two truth values {true(1), false(0)} as in classic propositional logic [**21**] [**22**].

Fuzzy If-Then Rules

Fuzzy if-then rules or *fuzzy* ***conditional statements*** [**20**] are expressions of the form ***IF A THEN B,*** where ***A*** and B are labels of ***fuzzy sets*** characterized by appropriate membership functions. Due to their concise form, fuzzy if-then rules are often employed to capture the imprecise modes of reasoning that play an essential role in the human ability to make decisions in an environment of uncertainty and imprecision. This can be described by the examples.

If the speed is ***high****, then apply the brake a* ***little.***

where speed and brake are linguistic variables, high and little are linguistic values or labels that are characterized by membership functions.

Another form of fuzzy if-then rule, proposed by Takagi and Sugeno, has fuzzy sets involved only in the premise part, while consequent part contains non fuzzy equations of input variable. By using Takagi and Sugeno's fuzzy if-then rule, we can describe the resistant force on a moving object as follows

IF velocity is high, THEN force = $\kappa * (velocity)^2$

where, again, ***high*** in the premise part is a linguistic label characterized by an appropriate membership function. Consequent part is non fuzzy equation of the input variable, velocity.

Both types of fuzzy if-then rules have been used extensively in both modeling and control. Through the use of linguistic labels and membership functions, a fuzzy if-then rule can easily capture the spirit of a "rule of thumb" used by humans. From another angle, due to the qualifiers on the premise parts, each fuzzy if-then rule can be viewed as a local description of the system under consideration. Fuzzy if-then rules form a core part of the fuzzy inference system which is discussed in next section.

Fuzzy Inference Systems (FIS)

Fuzzy inference systems are also known as fuzzy-rule-based systems, fuzzy models, fuzzy associative memories (FAM), or fuzzy controllers when used as controllers. Basically a fuzzy inference system is composed of five functional blocks as shown in Figure.4.1

- a ***rule base*** containing a number of fuzzy if-then rules;
- a ***database*** which defines the membership functions of
- a ***decision-making unit*** which performs the inference
- a ***fuzzification interface*** which transforms the crisp inputs
- a ***defuzzification interface*** which transform the fuzzy

Usually, the rule base and the database are jointly referred to as the ***knowledge base.***

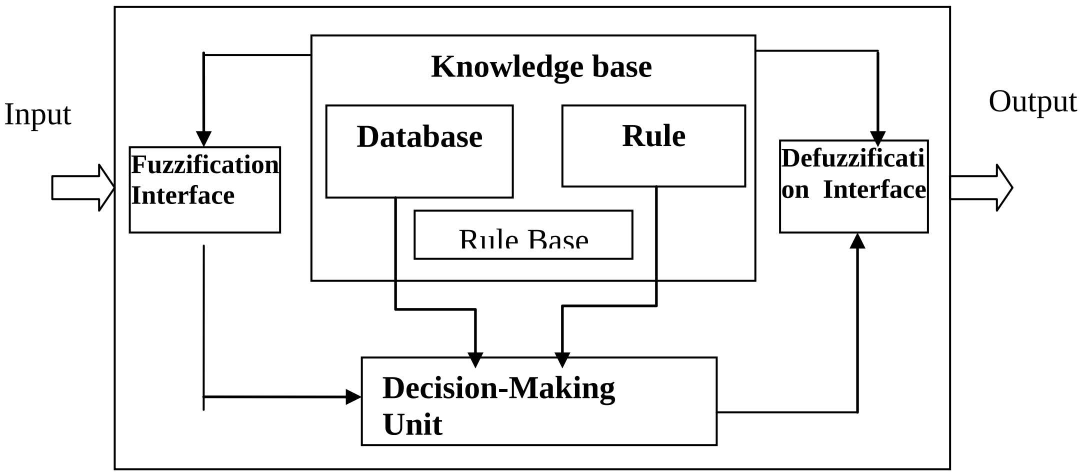

Figure 4. 1 Fuzzy Inference System.

The steps of **fuzzy reasoning** (inference operations upon fuzzy if-then rules) performed by fuzzy inference systems are:

1. Compare the input variables with the membership functions on the premise part to obtain the membership values (or compatibility measures) of each linguistic label. (This step is often called ***fuzzification)***.
2. Combine (through a specific T-norm operator, usually multiplication or min.) the membership values on the premise part to get ***firing strength (weight)*** of each rule.
3. Generate the qualified consequent (either fuzzy or crisp) of each rule depending on the firing strength.
4. Aggregate the qualified consequents to produce a crisp output. (This step is called ***defuzzification.)***

There are broadly three types of Fuzzy reasoning Models available in literature. They are Madmani fuzzy models, Sugeno fuzzy models (TSK model) and Tsukamoto fuzzy models. The first two models are most widely used. In ANFIS model Sugeno model is used widely because its rules are tunable based on input parameters.

4.2.2 ANFIS [23]

ANFIS is one of the Neuro-fuzzy model. Neural networks and fuzzy systems both are stand-alone systems. With the increase in the complexity of the process being modeled, the difficulty in developing dependable fuzzy rules and membership functions increases. This has led to the development of another approach which is mostly known as ANFIS approach. It has the benefits of both neural networks and fuzzy logic. One of the advantages of fuzzy systems is that they describe fuzzy rules, which fit the description of real-world processes to a greater extent. Another advantage of fuzzy systems is their interpretability; it means that it is possible to explain why a particular value appeared at the output of a fuzzy system. In turn, some of the main disadvantages of fuzzy systems are that expert's knowledge or instructions are needed in order to define fuzzy rules, and that the process of tuning of the parameters of the fuzzy system (e.g. parameters of the membership functions) often requires a relatively long time. Both these disadvantages are related to the fact that it is not possible to train fuzzy systems. A diametrically opposite situation can be observed in the field of neural networks. It is known that neural networks are trained, but it is extremely difficult to use a prior knowledge about the considered system and it is almost impossible to explain the behavior of the neural network system in a particular situation. In order to compensate the disadvantages of one system with the advantages of another system, several researchers tried to combine fuzzy systems with neural networks. A hybrid system named ANFIS (Adaptive-Network-Based Fuzzy Inference System) has been proposed by Jang [**20**]. Fuzzy inference in this system is realized with the aid of a training algorithm, which enables to tune the parameters of the fuzzy system.

ANFIS ARCHITECTURE

The ANFIS is a fuzzy Sugeno model put in the framework of adaptive systems to facilitate learning and adaptation. The Sugeno fuzzy model was proposed by Takagi & Sugeno in an effort to formalize a systematic approach to generating fuzzy rules from an input-output data set.

A typical fuzzy rule in a Sugeno fuzzy model has the format

IF x is A and y is B THEN z = f(x,y),

where A and B are fuzzy sets in the antecedent; z = f(x, y) is a crisp function in the consequent.

Usually f(x, y) is a polynomial in the input variables x and y, but it can be any other functions that can appropriately describe the output of the system within the fuzzy region specified by the antecedent of the rule. If f(x, y) is a first-order polynomial, than model is called as the first-order Sugeno fuzzy model. If f is a constant, then it is called the zero-order Sugeno fuzzy model, which can be viewed either as a special case of the Mamdani fuzzy inference system, where each rule's consequent is specified by a fuzzy singleton, or a special case of Tsukamoto's fuzzy model where each rule's consequent is specified by a membership function of a step function centered at the constant. Moreover, a zero order Sugeno fuzzy model is functionally equivalent to a radial basis function network under certain minor constraints [**20**].

Considering a first-order Sugeno fuzzy inference system which contains two rules:

Rule 1: **IF** X is A_1 AND Y is B_1, THEN
$f_1 = p1\,x + q_1 y + r_1$

Rule 2: **IF** X is A_2 AND Y is B_2, **THEN**
$f_2 = p_2\,x + q_2\,y + r_2$.

Figure 4.2 illustrates graphically the fuzzy reasoning mechanism to derive an output f from a given input vector [x, y]. The firing strengths w_1 and w_2 are usually obtained as the product of the membership grades in the premise part, and the output f is the weighted average of each rule's output.

To facilitate the learning of the Sugeno fuzzy model, it is convenient to put the fuzzy model into framework of adaptive networks that can compute gradient vectors systematically. The resultant network architecture, is ANFIS, that is shown in Figure 4.3, where node within the same layer performs functions of the same type, as detailed below. Here circle indicates a fixed node, whereas a square indicates an adaptive node.

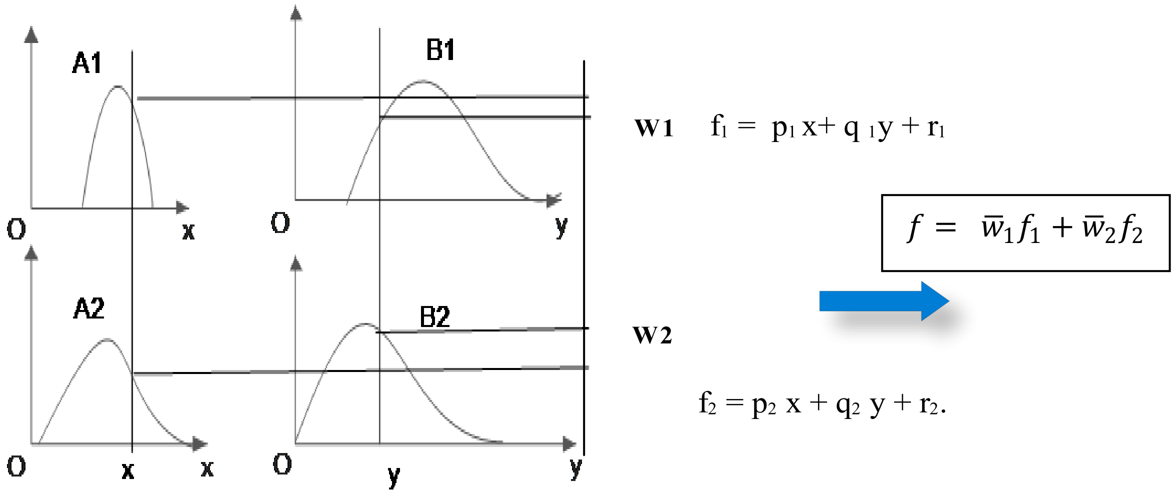

Figure 4. 2 First order sugeno model

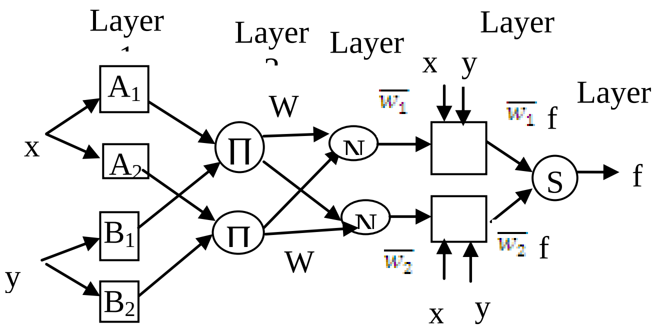

Figure 4. 3 ANFIS architecture.

Layer 1 Each node in this layer generates membership grades of a linguistic label. For instance, the node function of the i−th node may be a generalized bell membership function:

$$O^1_i = \mu_{Ai}(x) = \frac{1}{1+((x-\frac{c_i}{a_i})^2)^{bi}} \qquad i=1,2 \qquad (4.1)$$

where x is the input to node i; A_i is the linguistic label (small, large, etc.) associated with this node; and {ai, bi, ci} is the parameter set that changes the shapes of the membership function. Parameters in this layer are referred to as the premise parameters.

Layer 2 Each node in this layer calculates the firing strength of a rule via multiplication and the nodes are fixed:

$$O_i^2 = w_i = \mu_{Ai}(x)\, \mu_{Bi}(y), \qquad i=1,2 \qquad (4.2)$$

Layer 3 The nodes are fixed nodes. They are labeled with *N*, indicating that they play a normalization role to the firing strengths from the previous layer. The outputs of this layer can be represented as:

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1+w_2}, \qquad i = 1, 2 \qquad (4.3)$$

Layer 4 The nodes are adaptive nodes. The output of each node in this layer is simply the product of the normalized firing strength and a first-order polynomial (for a first-order Sugeno model). Thus, the outputs of this layer are given by:

$$O^4_i = \overline{w}_i f_i = \overline{w}_i(p_i x + q_i y + r_i) \qquad i = 1,2 \qquad (4.4)$$

where $\overline{w}$ is the output of layer 3, and $\{p_i, q_i, r_i\}$ is the parameter set. Parameters in this layer will be referred to as the consequent parameters.

Layer 5 There is only one single fixed node labeled with *S*. This node performs the summation of all incoming signals. Hence, the overall output of the model is given by:

$$O^5_i = \sum_{i=1}^{2} \overline{w}_i f_i = \frac{(\sum_{i=1}^{2} \overline{w}_i f_i)}{w_1+w_2} \qquad (4.5)$$

The shown adaptive network in Figure 4.3 is functionally equivalent to a fuzzy inference system in Figure 4.3. The basic learning rule of ANFIS is the back propagation gradient descent [**23**], which calculates error signals (the derivative of the squared error with respect to each node's output) recursively from the output layer backward to the input nodes. This learning rule is exactly the same as the back propagation learning rule used in the common feed forward neural networks. From the ANFIS architecture in Figure 4.3, it is observed that given the values of premise parameters, the overall output f can be expressed as linear combinations of the consequent parameters:

$$f = \overline{w}_1 f_1 + \overline{w}_2 f_2 \qquad (4.6)$$

$$f = \overline{w}_1(p_1x+q_1y+r_1) + \overline{w}_2(p_2x+q_2y+r_2) \qquad (4.7)$$

4.2.3 ANFIS Learning Method

Neural networks, the back propagation algorithm are used to learn, or adjust weights on connecting arrows between neurons from input-output training samples. In the ANFIS structure, the parameters of the premises and consequents play the role of weights. Specifically, the shape of membership functions in the "If" part of the rules is determined by a finite number of parameters. These parameters are called premise parameters, whereas the parameters in the "THEN" part of the rules are referred to as consequent parameters. The ANFIS learning algorithm(Jang [**20**]) consists of adjusting the above set of parameters.

For ANFIS, a mixture of back propagation and least square estimation (LSE) is used. Back propagation is used to learn the premise parameters, and LSE is used to determine the parameters in the rules' consequents. A step in the learning procedure has two passes. In the forward pass, node outputs go forward, and the consequent parameters {pi, qi, ri} are estimated by least squares method, while the premise parameters remain fixed. In the backward pass the error signals are propagated backwards, and back propagation is used to modify the premise parameters {ai, bi, ci}, while consequent parameters remain fixed. This combination of least-squares and back propagation methods are used for training FIS membership function parameters to model a given set of input/output data. The performance of this system will be evaluated using RMSE, root mean square errors (difference between the FIS output and the training/testing data output), defined as:

$$\text{RMSE} = \frac{1}{n}\sqrt{\sum_{k=1}^{n}(y_k - o_k)^2} \tag{4.8}$$

where y_k is the desired output and o_k is the actual system output. n is the number of training/testing samples. In thesis RMSE used for comparison of performance of ANFIS and NN method.

4.2.4 Membership Function and Rules Selection for ANFIS

In a conventional fuzzy inference system, the number of rules is decided by an expert who is familiar with the target system to be modeled. In ANFIS simulation, however, no expert is available and the number of membership functions (MF's) assigned to each input variable is chosen empirically, that is, by plotting the data sets and examining them visually, or simply by trial and error. For data sets with more than three inputs, visualization techniques are not very effective and most of the time we have to rely on trial and error. This situation is similar to that of neural networks; there is just no simple way to determine in advance the minimal number of hidden units needed to achieve a desired performance level. There are several other techniques for determining the numbers of MFs and rules, such as

CART and clustering methods. In a fuzzy inference system, basically there are three types of input space partitioning:

- Grid partitioning method
- Scatter partitioning method :It includes
    - Fuzzy-C means clustering method.
    - Subtractive clustering method.
- Tree Partitioning method.

In thesis FIS is generated by using Grid partitioning and scatter partitioning. Both methods are explained in subsection.

Grid petitioning

Grid partitioning is an approach for initializing the structure in a fuzzy inference system. In this method it generates rules by enumerating all possible combinations of membership functions of all inputs. The number of MFs on each input variable uniquely determines the number of rules. The initial values of premise parameters are set in such a way that the centers of the MFs are equally spaced along the range of each input variable. Figure 4.4 shows example of grid portioning.

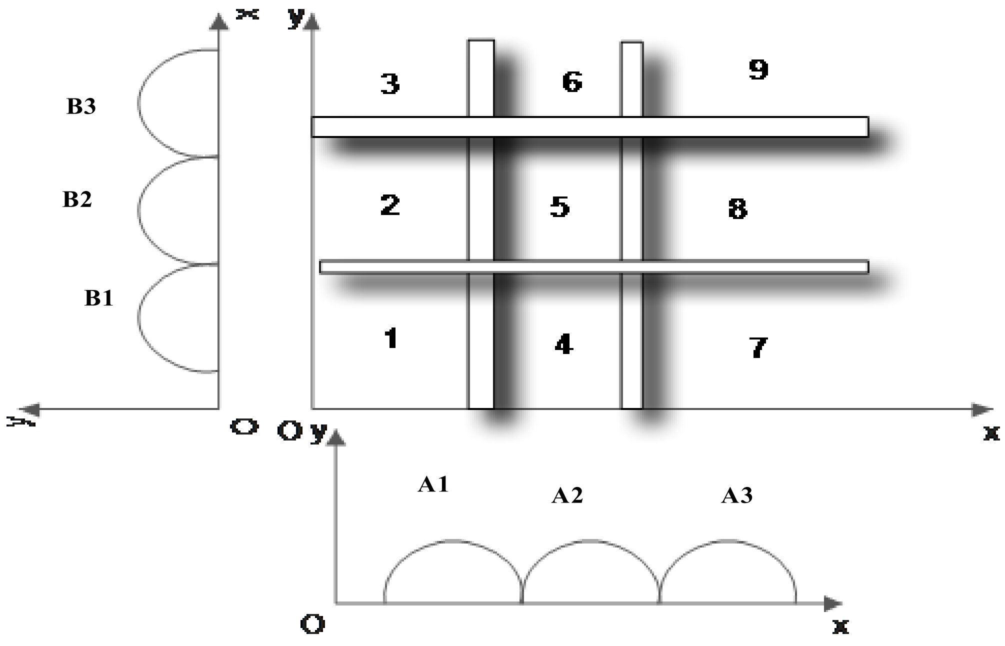

Figure 4. 4 Grid partitioning of input space for two input sugeno fuzzy model with nine rules .

The grid-partitioning approach to fuzzy systems has the serious disadvantage that the very regular partition of the input space may be unable to produce a rule set of acceptable size which is able to handle a given data set well. If, for example, the data contains regions with several small clusters of different classes, then small rule patches have to be created to classify the data in this region correctly. This problem becomes even more serious as the dimension of the input data increases. That leads to an exponential explosion. For instance, for a fuzzy inference system with 10 inputs, each with two membership functions, the grid partitioning leads to 1024 (=2^10) rules, which is inhibitively large for any practical learning methods. The "curse of dimensionality" refers to such situation where the number of fuzzy rules, when the grid partitioning is used, increases exponentially with the number of input variables. So this leads to increase of simulation time and poor results for high dimensional problem. So to overcome this Scatter partitioning methods are used.

Scatter partitioning

To eliminate the problems associated with grid-partitioning, other ways of dividing the input space into rule patches have been proposed. That approach, known as [**24**] *scatter partitioning*, is to allow the IF-parts of the fuzzy rules to be positioned at arbitrary locations in input space. If the rules are represented by *n*-dimensional Gaussians or normalized Gaussians , this means that the centers of the Gaussians are not anymore confined to corners of a rectangular grid. Rather, they can be chosen freely, e.g., by a clustering algorithm working on the training data. So two clustering algorithms have been used as mentioned previously: 1) Fuzzy C means clustering and 2) subtractive clustering

Fuzzy C-means clustering (FCM):

Clustering partitions a data set into several groups such that the similarity within a group is larger than that among groups. Achieving such a partitioning re- requires a similarity metrics that takes two input vectors and returns a value reflecting their similarity. Since most similarity metrics are sensitive to the ranges of elements in the input vectors, each of the input variables must be normalized to within, say, the unit interval [0, 1].

Fuzzy C-means clustering (FCM), also known as fuzzy ISODATA, is a data clustering algorithm in which each data point belongs to a cluster to a degree specified by a membership grade. Bezdek proposed this algorithm in 1973. FCM partitions a collection of n vector x_i i = 1, ..., n into c fuzzy groups, and finds a cluster center in each group such that a cost function of dissimilarity measure is minimized. FCM employs fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by membership grades between 0 and 1. Here number cluster center represents the number rules. Here number of rules can be fixed by us.

Subtractive Clustering (SC):

When there is no clear idea how many clusters there should be for a given set of data, *subtractive clustering* is a fast, one-pass algorithm for estimating the number of clusters and cluster centers in a set of data. [**25**]Subtractive clustering operates by finding the optimal data point to be defined as a cluster center, based on the density of surrounding data points. All data points within the radius distance of these points are then removed, in order to determine the next data cluster and its center. This process is repeated until all of the data is within the radius distance of a cluster center. This method used for rules generation when number of inputs larger. It gives optimized rules by taking into radii specified. In this work all the three types FIS method are used and compared for data prediction

4.3 ANFIS based data rate prediction: Basic Scheme

4.3.1 Preparation procedure:

ANFIS model was considered to be tuned in an arbitrary radio configuration e.g. IEEE WLAN 802.11g as discussed in previous chapter 3. ANFIS has been trained with back propagation algorithm . For given set of input in basic scheme, target data rate, r_k^{tgt} is found by Eq.3.12. Basic scheme ANFIS model is depicted in Figure 4.5. Three types ANFIS techniques were considered. They are ANFIS structure using the grid portioning, FCM based ANFIS technique and subtractive clustering based ANFIS technique. For smaller dimensional data grid partitioned based FIS structure can be used, here number of membership functions can decided by user. As data input increases more than five, problem of "*curse of dimensionality"* arises. Hence this method is suited only for small number of input.

Second case FCM based FIS structure is used for prediction number of rules can be fixed by fixing fuzzy center but membership functions cannot be fixed. Membership functions and rules are generated by unsupervised FCM method. Subtractive FIS method is one that can be used to train ANFIS fast for high dimensional problem. Based radii of influence selected fuzzy rules are generated on its own and here also membership functions generated by unsupervised method.

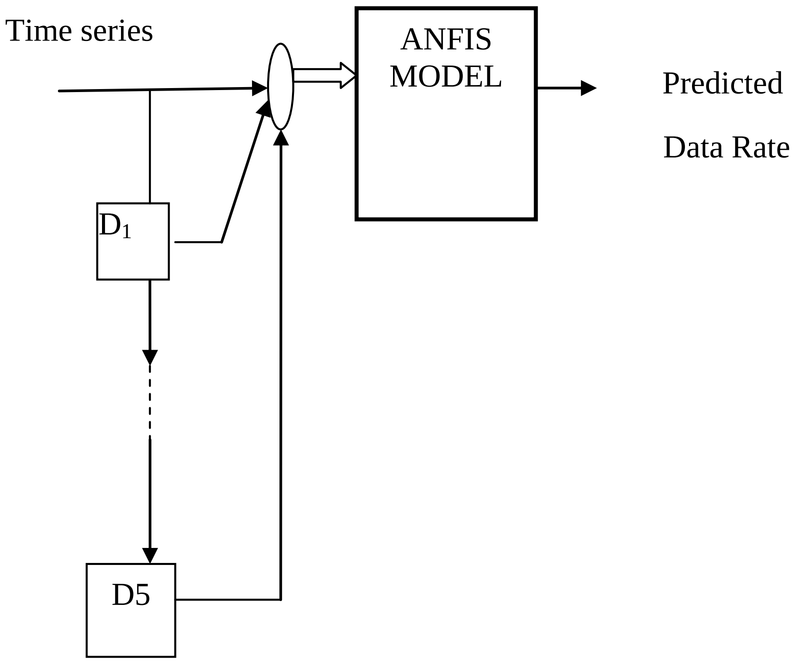

Figure 4. 5 ANFIS model for simulation- Basic Scheme

Users don't have choice fixing MFs once subtractive clustering is used. But It shown that this method works better and gives good result in terms Data rate prediction accuracy for problems involving higher number of inputs. It gives optimized rules, which in turn reduces ANFIS tuning complexity. As mentioned previously learning model has to go into the SDR so ANFIS can be best suited for it .

4.3.2 Simulation results and discussion

Similar to neural network method used here, ANFIS was considered, tuned to specific, arbitrary radio configuration, e.g. WLAN 802.11g and reference bit rate values have been set to $|M| = 6$. It is assumed that the radio scene analysis phase has generated time series of data rate according following probability distribution [6, 12, 24, 36, 48, 54: .5 .2 .1 .1 .07 .03]. Here bigger probability is assigned to the appearance of m_1=6. Time window n is set to 5 and smoothing factor a of exponentially moving average algorithm is set to .362 accordingly weights for time window are β_i={ 0.2310 0.1473 0.0940 0.0600 0.0383}. Here performance measure in terms RMSE and prediction accuracy were used as performance idex. ANFIS based results are compared with reference work of NNs, which was simulated in previous chapter.

In first case conventional ANFIS was considered, which uses grid partitioning method to generate rules. In this five inputs are given as input to ANFIS and 2 Gaussian shaped membership functions are taken for each input. Accordingly it has generated 2^5 =32 rules. Gaussian shapes were chosen first because nonlinear parameters to be tuned are only two. The membership is given by following equation:

$$F(x,\sigma,c) = e^{\frac{-(x-c)^2}{2\sigma^2}} \tag{4.9}$$

Here σ and c have to be tuned so for 5 input case there are 20 nonlinear parameters to be tuned. Since there are 32 rules so total linear parameters were calculated based on consequent side equation of 4.10

$$f_1 = p_1 a_1 + q_1 b_1 + r_1 c_1 + s_1 d_1 + t_1 e_1 + g_1 \tag{4.10}$$

If a, b, c, d and e are five inputs to the ANFIS than there are 6 linear parameters to vary including constant. Thus total number linear parameters were 6*32=192 and total parameters to be tuned are 20+192=212. To achieve good generalization capability, it is important that the number of training data points be several times larger than the number parameters being estimated so 1000 data points were taken for training. As in NNs technique testing is done with seen data and unseen data for 100 data points each. Here error performance measured with RMSE, which is mentioned previously. Simulation was conducted for 500 epochs. The conventional ANFIS membership function before training and after training, are presented in Figure 4.6 and 4.7.RMSE plot for training and validation case is shown in Figure 4.8, whereas Figure 4.9 and 4.10 shows prediction accuracy in case of training and validation case. RMSE, Prediction accuracy parameters tuned number of rules used are tabulated in Table 4.1.

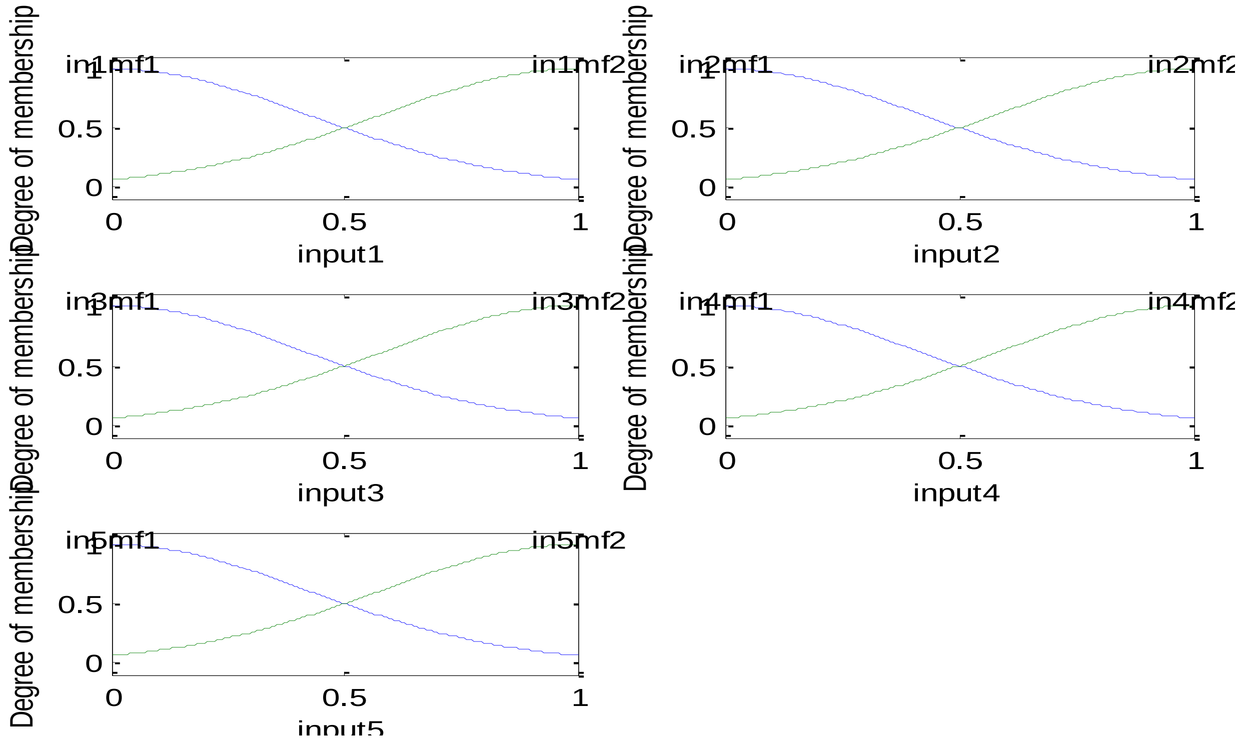

Figure 4. 6 Memberships plot for each input before training.

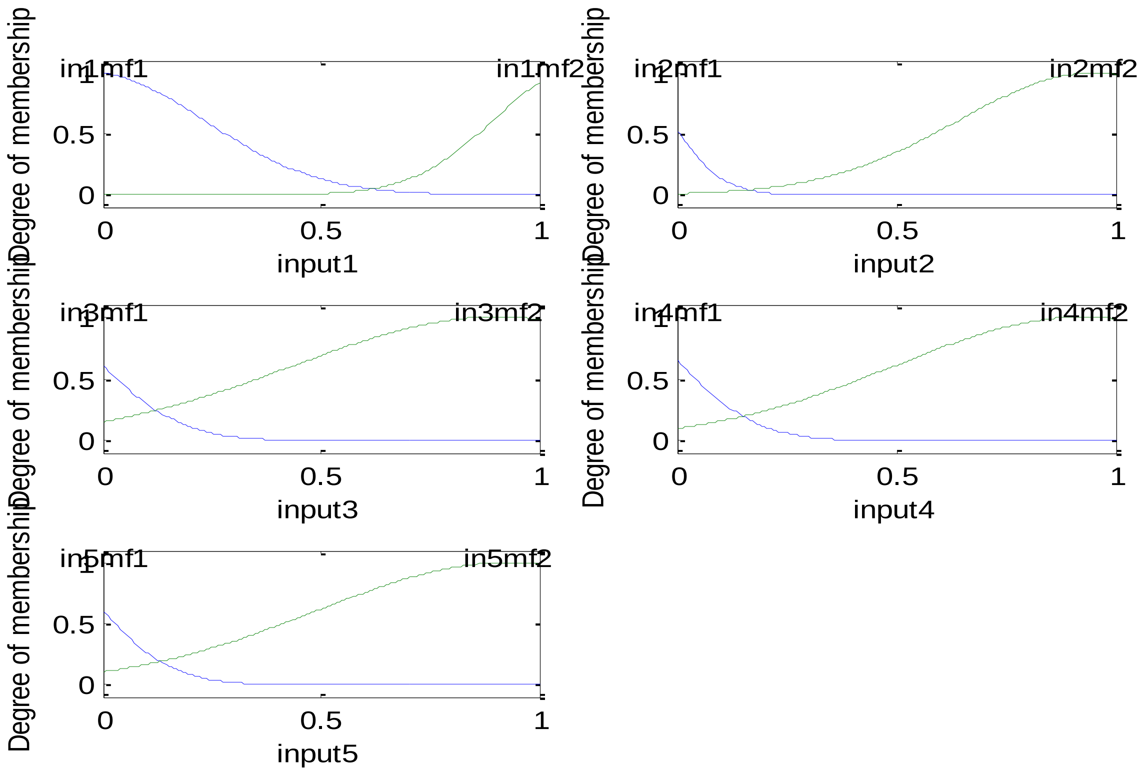

Figure 4. 7 Membership plot after training

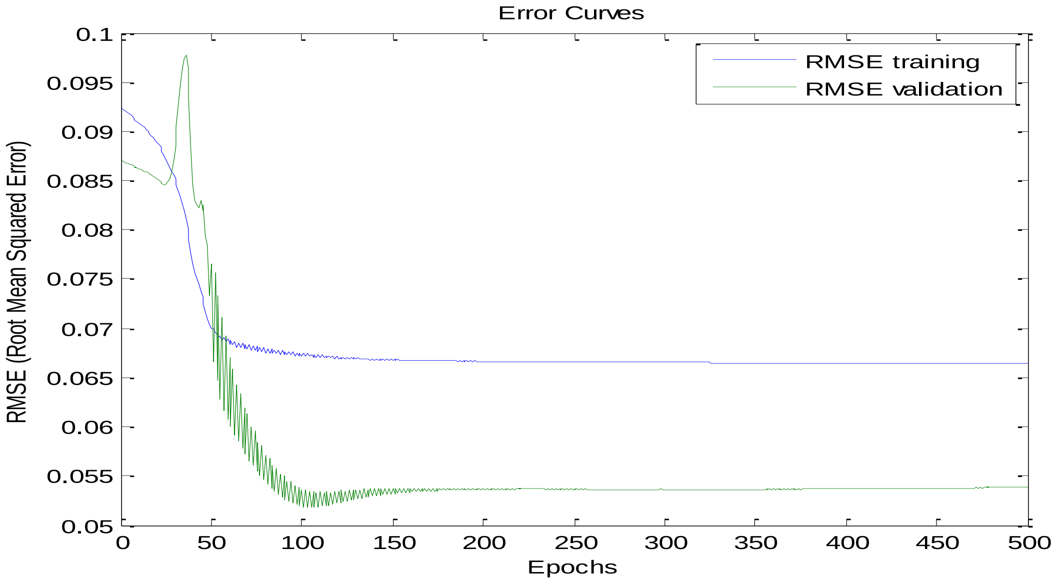

Figure 4. 8 RMSE for training and validation.

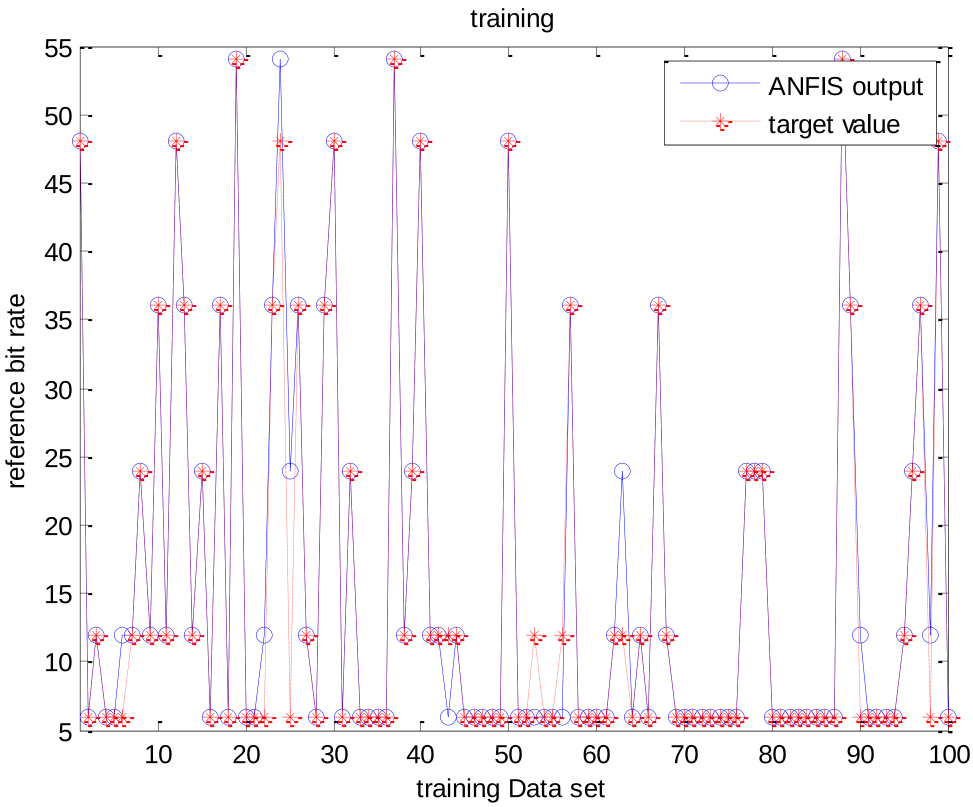

Figure 4. 9 Prediction accuracy of conventional ANFIS in training sequence-basic scheme.

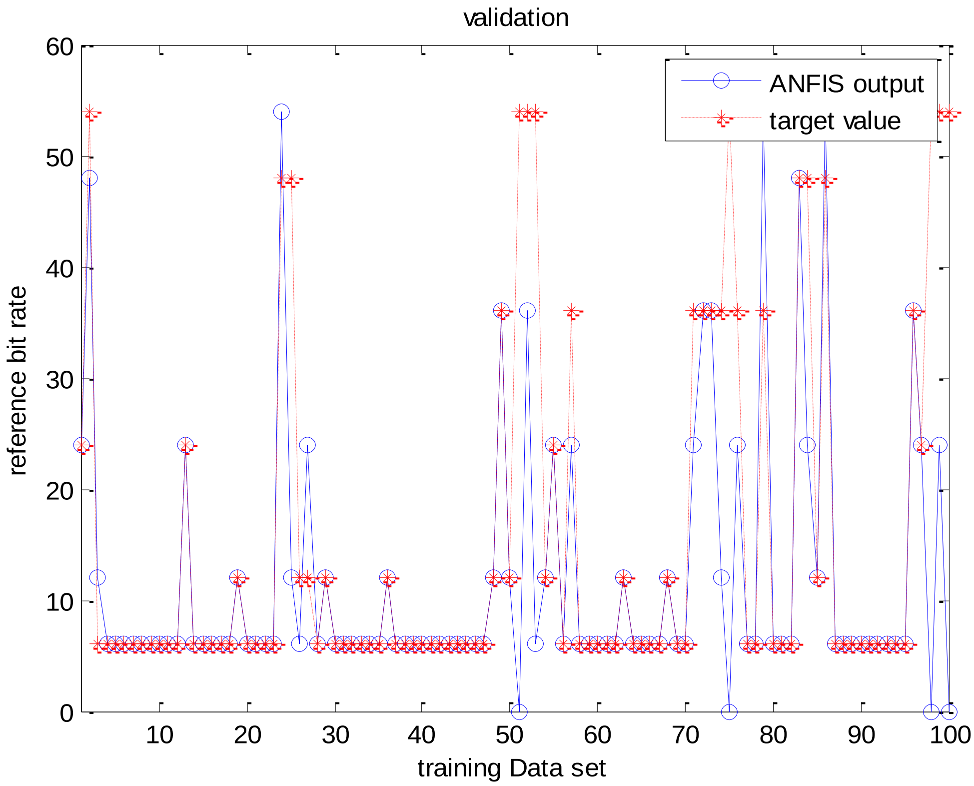

Figure 4. 10 Prediction accuracy of conventional ANFIS in validation sequence -basic scheme

Figure.4.7 presents tuned membership functions after training. Figure.4.9 and Figure 4.10 depicts that prediction accuracy of conventional ANFIS is 91% during training and 89 % in validation. Whereas Elman network prediction accuracy is 83 % during training and 81 % during validation. Thus from this it could be concluded that ANFIS has better accuracy than ENN. Figure 4.8 depicts RMSE curves for ANFIS prediction. From Table 4.1 it is observed RMSE error is more in case of ENN as compared to ANFIS.

FCM based ANFIS method was tested next. Here rules are predetermined by fixing number of centers. As mentioned previously it generates FIS structure by scatter partitioning. Here membership functions were assigned automatically by software. So number of tuning parameters is reduced in this case by reducing number of rules. Hence all simulation conditions remained same as previous methods. The results for optimum cluster size are presented with best prediction accuracy and RMSE in case of training and validation. Figure 4.11 -4.15 present results of simulation. The summary is tabulated at Table 4.1. In first trial 20 rules were taken which gives 248 parameters. This took long time. Hence 15 rules were taken

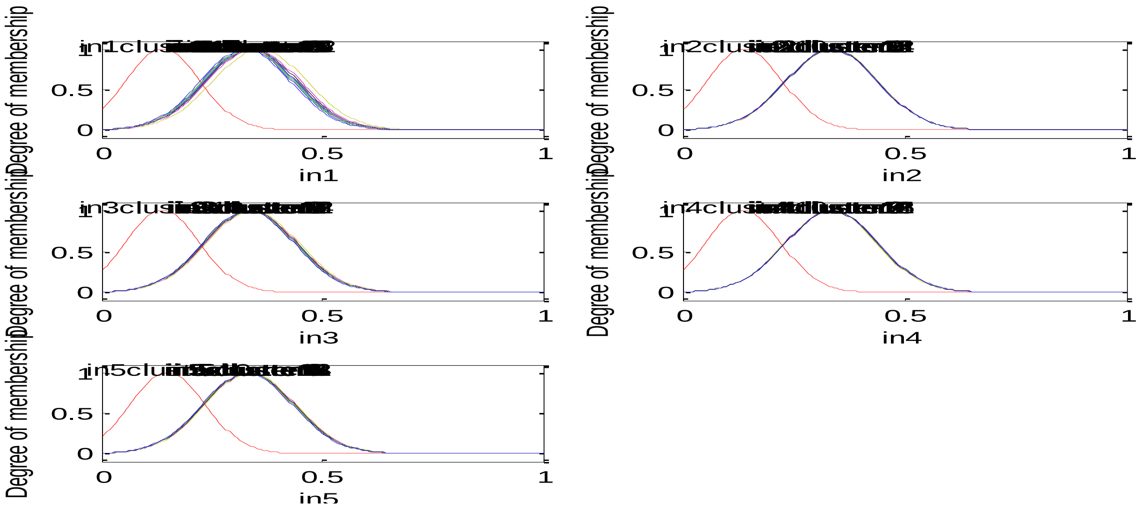

Figure 4. 11 Memberships plot for each input before training in case of FCM based structure.

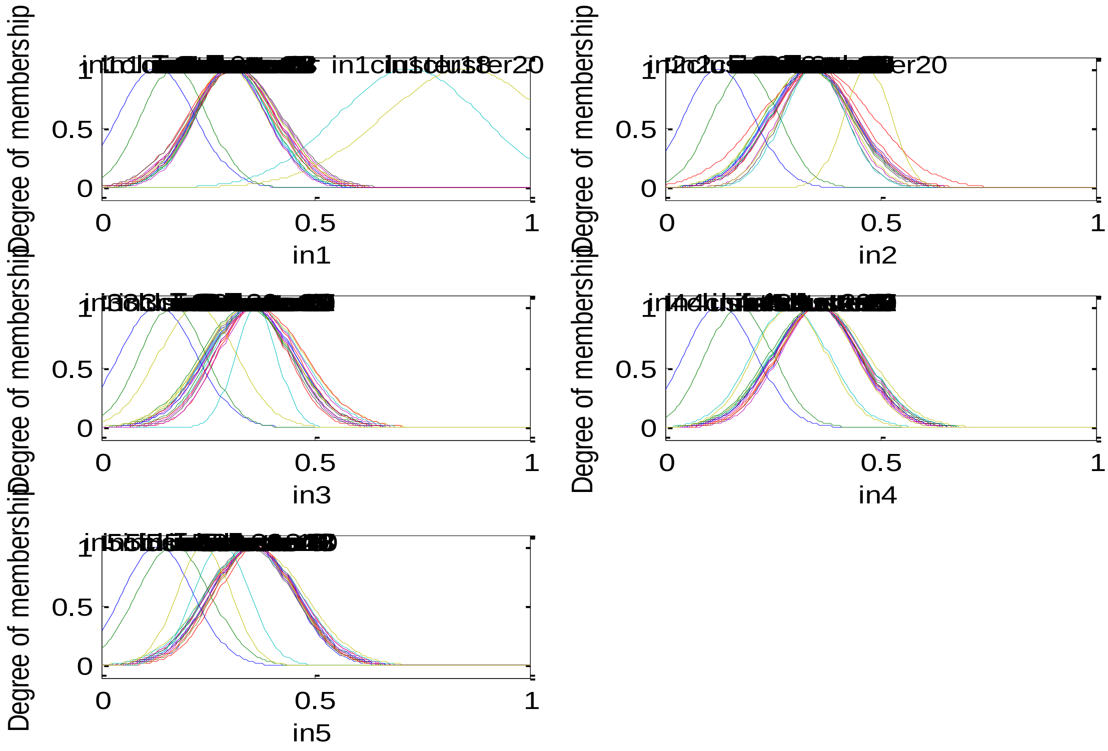

Figure 4. 12 Memberships plot for each input after training in case of FCM based structure.

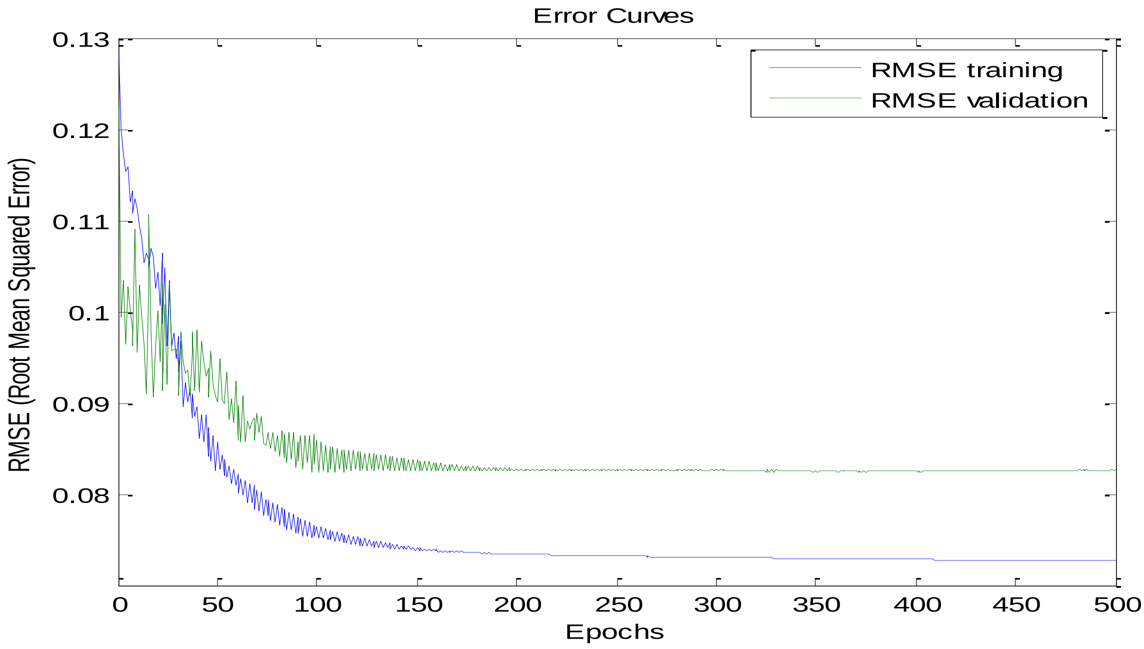

Figure 4. 13 RMSE for training and validation in case FCM based ANFIS.

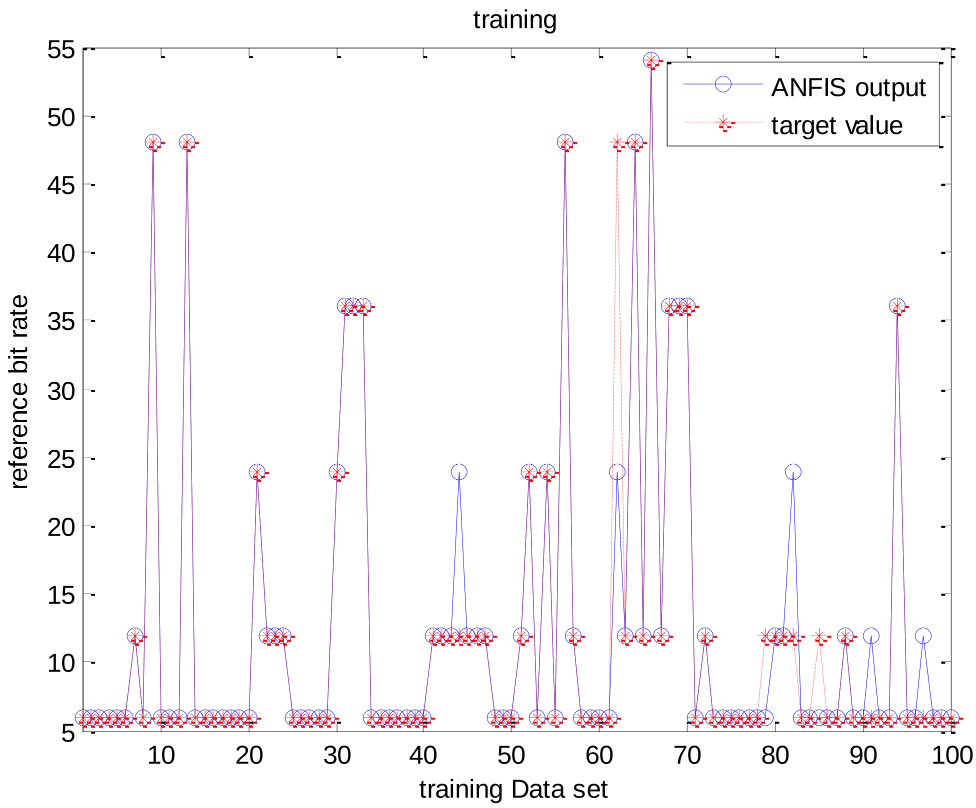

Figure 4. 14 Prediction accuracy of FCM based ANFIS in training sequence-basic scheme.

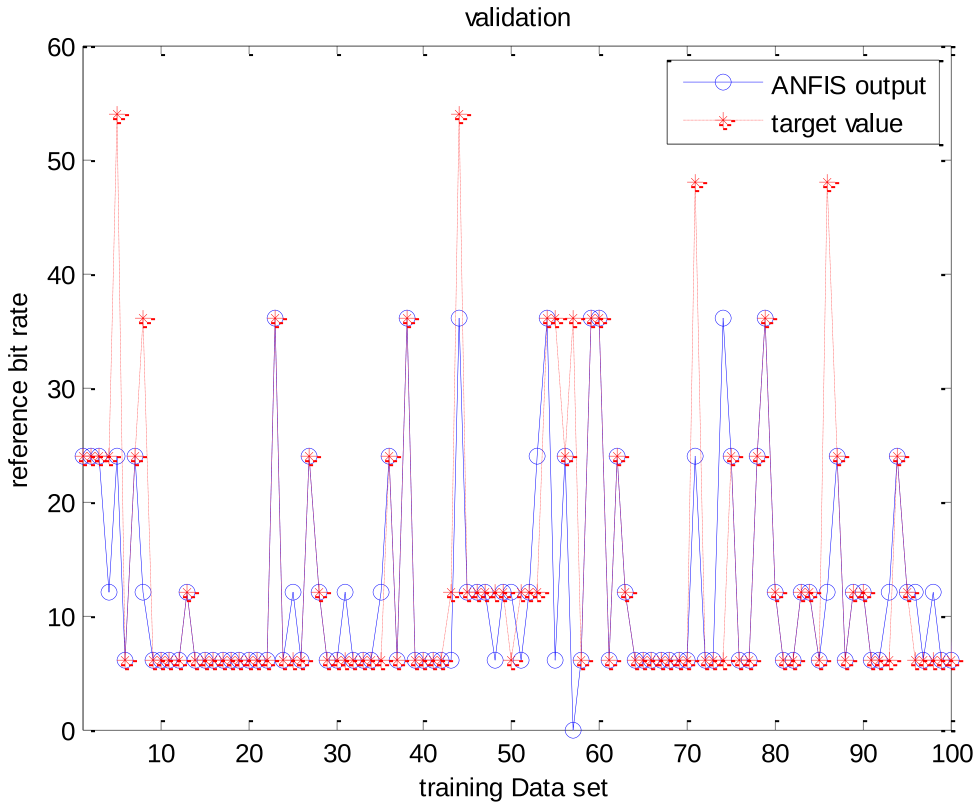

Figure 4. 15 Prediction accuracy of FCM based ANFIS in validation sequence-basic scheme

Last ANFIS with subtractive clustering method was used generate FIS structure. As discussed previously, when there is no idea of how many clusters to be selected, than this method is one of a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a set of data. Here cluster radius has to be mentioned, the cluster radius indicates the range of influence of a cluster when you consider the data space as a unit hypercube. Specifying a small cluster radius usually yields many small clusters in the data, and results in many rules. Specifying a large cluster radius usually yields a few large clusters in the data, and results in fewer rules. An important advantage of using a clustering method to find rules is that the resultant rules are more tailored to the input data than they are in a FIS generated without clustering. This reduces the problem of an excessive propagation of rules when the input data has a high dimension. Here simulation done with taking different cluster radius. Results of best structure are presented. Here also 1000 data points are used for training and simulation is run for 500 epochs. Figure 4.16 to 4.120 presents all simulation results with radius influence kept as .5 .This method fast as compared to conventional and FCM based method. Figure 4.16 and Figure 4.17 present membership function before training and after training. Figure.4.19 and Figure.4.20 depicts prediction accuracy with training data set and validation data set. These depicts that prediction accuracy for case training set 91 % and with validation set 86 %. When compared to ENN prediction accuracy is more better. Figure.4.18 presents RMSE curve in case training and validation data set.

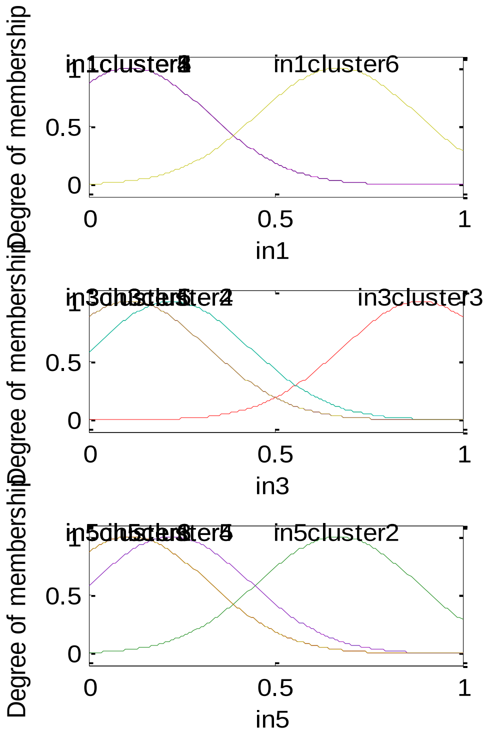

Figure 4. 16 Memberships plot for each input before training in case of subtractive clustering based structure.

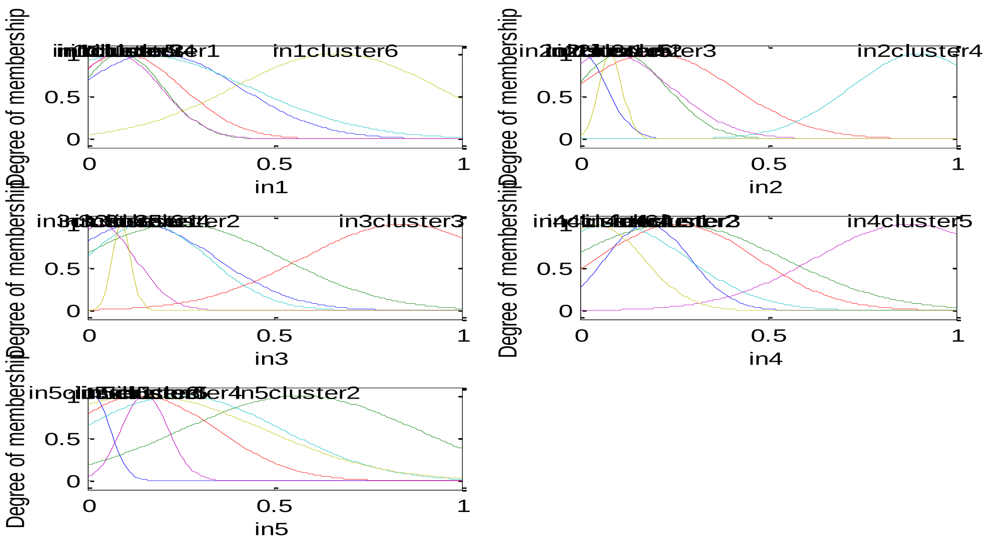

Figure 4. 17 Memberships plot for each input after training in case of subtractive clustering based structure.

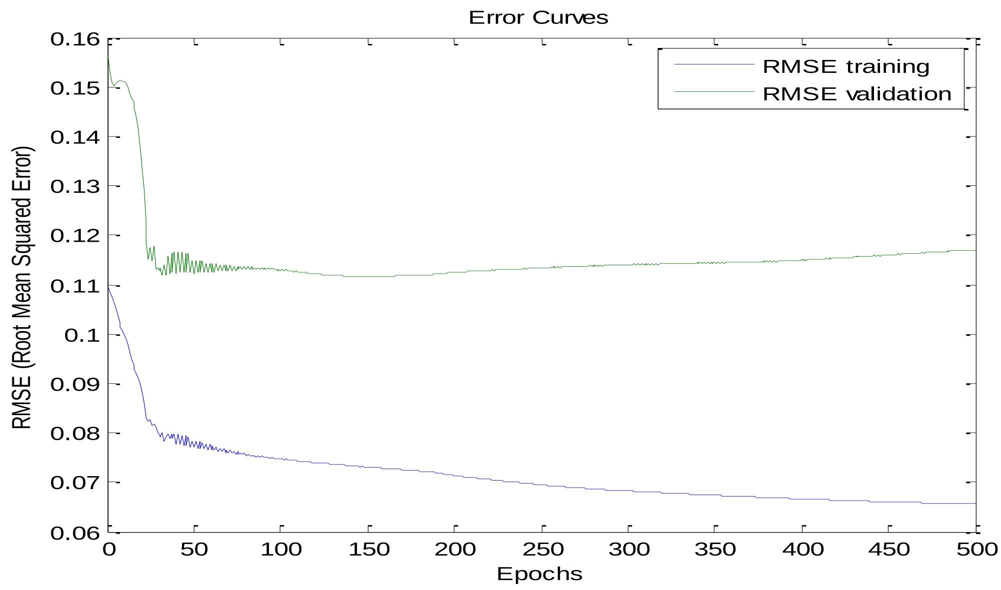

Figure 4. 18 RMSE for training and validation in case subtractive clustering based ANFIS.

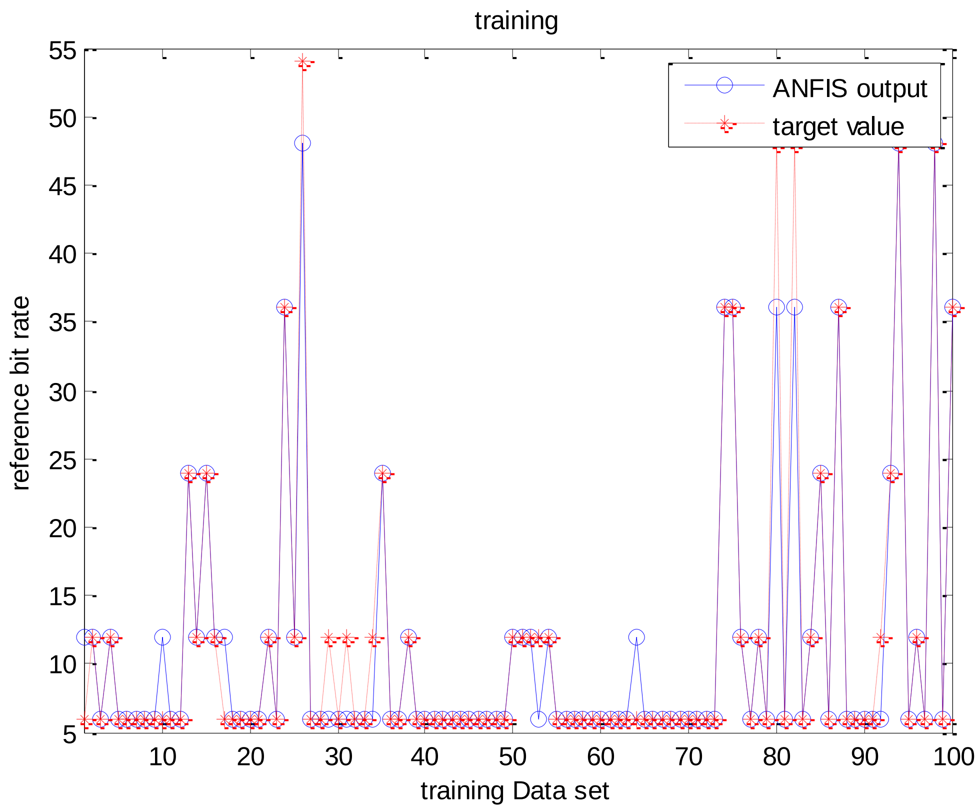

Figure 4. 19 Prediction accuracy of subtractive clustering based ANFIS in training sequence-basic scheme

\

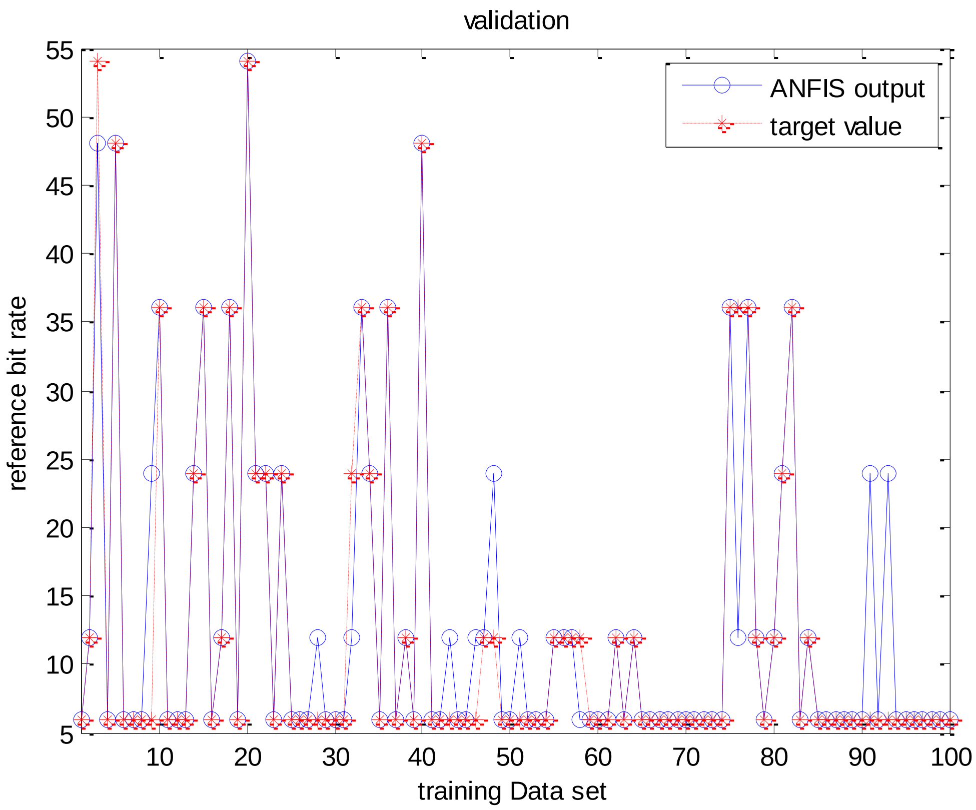

Figure 4. 20 Prediction accuracy of subtractive clustering based ANFIS in validation sequence-basic scheme

In Table 4.1 all simulation parameters have been tabulated including best case of Elman neural network for comparison. All results are for basic scheme. It can be observed from the table that conventional ANFIs has good prediction accuracy and difference between $RMSE_{train}$ and $RMSE_{validation}$ very less compared to other networks. But total tunable parameters high in this case 20 nonlinear parameters and 192 linear parameters. So it takes longer time to train. If the number input is increased it faces problem of "*curse of dimensionality*". So cluster based algorithm are used to increase speed execution. From Table 4.1 FCM based FIS generation gives best result in case of 20 clusters or rules used, but it also generates huge tunable parameters. So it also faces same problem as previous and here optimum rules cannot be fixed, same trial and error method must be used fix the rules.

Last test case with subtractive clustering based ANFIS provided best results are found when radius of influence kept as .5. As said here optimized rules are generated by FIS only. So it generates less number of rules and gives good accuracy. Total number of tunable parameters very less i.e. 30+50= 80. So it helps to speed up learning algorithm when dimension problem is increased. For future studies only subtractive clustering was used.

Table 4.1 Performance index of all ANFIS techniques-basic scheme

Type of techniques used	No of Hidden nodes	Number of rules or centers	Linear parameters	Nonlinear parametrs	RMSE train	RMSE validation	RMSE train-RMSEvalid	Prediction accuracy training	Prediction accuracy validation
ANFIS grid partition	92	32	192	20	0.0518	0.0575	0.0057	91	89
ANFIS-FCM	128	10	60	100	0.0823	0.0885	0.0062	83	77
ANFIS-FCM	188	15	90	150	0.0599	0.0971	0.0372	91	83
ANFIS-FCM	248	20	120	200	0.0316	0.069	0.0374	98	91
ANFIS-SC-.3 radius	116	9	54	90	0.0841	0.1101	0.026	85	85
ANFIS-SC-.4	68	5	30	60	0.099	0.1117	0.0127	85	85
ANFIS-SC-.5	68	5	30	50	0.0547	0.0872	0.0325	91	86
ANFIS-SC-.6	80	6	36	60	0.0766	0.1011	0.0245	89	84
Elman neural network	15		15*15+5*15+15+5=335		0.1161	0.102	0.0141	83	81

Comparing with neural network based technique ANFIS based technique out performs in all performance parameters. As tabulated in table 4.1 total numbers weights to be trained are huge in case recurrent Elman network with 15 hidden nodes. Total 335 weights have to be updated and there are 75 non-linear functions have to be solved. In case of conventional ANFIS if two membership functions are considered for each input, only 10 nonlinear functions need to be solved. And 32 rules lead to 32 linear equations. So total equations to be solved simple is less. From above table it can be seen that RMSE error between validation and training for the case of conventional ANFIS, best case of FCM based ANFIS and SC-ANFIS are .0057,.0062 and .0121 respectively. Where in neural network method it gave error of 0.0141 which is more than ANFIS basd method and prediction accuracy also more than neural network method. Thus ANFIS method provides better performance than NN methods.

4.4 ANFIS based data rate prediction: Extended Scheme

4.4.1 Preparation procedure:

As discussed in previous chapter, the complexity of problem is extended by further taking into account of time –zone parameter. Similarly day divided into four time zones and during each of them, the configuration in question is associated with mean, most usually observed data rate value, which is denoted as $\overline{m}_{tz} \in M$.As mentioned previous chapter mean value enhances the learning scheme with feature of past experience. Here neural network method is used as reference model to compare. As similar to basic scheme time window of n=8 slots was considered. Here complexity of the problem is increased so window length was also increased. Detailed procedure mentioned in previous chapter subsection 3.6.1 is used prepare time series of data rate and $r_k^{tgt,ext}$. Here smoothing factor for time window $\chi = .7$ accordingly weights β_i' change . Here also it was assumed that the day is divided in four equal time zones as follows: 06:00–12:00, 12:00–18:00, 18:00–24:00 and 00:00–06:00. In each of these time zones a different mean value $\overline{m}_{tz}$ is observed, let them be set equal to 24, 6, 36 and 48 Mbps for each of the four time zones, respectively; this might reveal for instance the existence of high load situation during the mid working day.

Since dimension of the problem is increased to 8, so conventional ANFIS is not used predict the data rate, because it generates huge rules which cannot be handled by the simulation environment. FCM based method also not used because here we need to do trial error method to get optimum rules which is time consuming. Only faster method based on subtractive clustering was used to generate FIS. SC-based ANFIS was used for prediction, because it has ability remove redundant rules and give best fit rules which determine the prediction. ANIFS model for training is shown in Figure 4.21

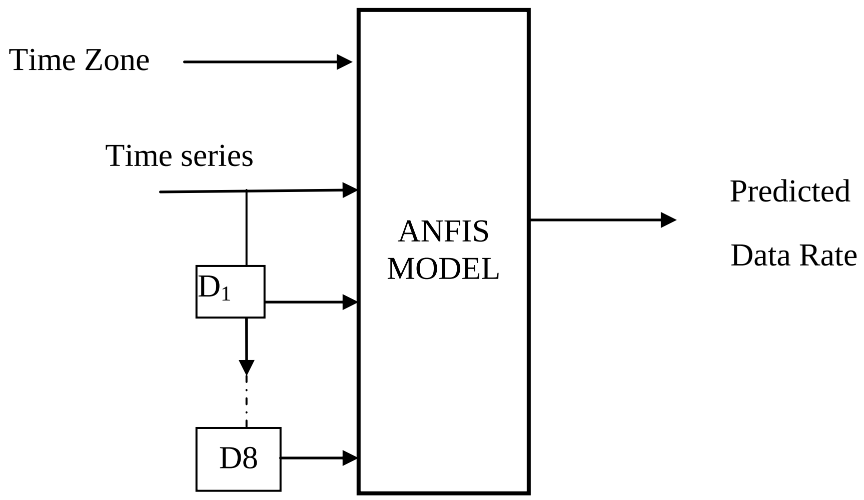

Figure 4. 21 ANFIS model for the extended learning scheme

4.4.2 Simulation results and discussion:

RMSE, Prediction accuracy were used as performance index. Here while comparing with neural network of extended case, Mathematical complexity is also considered for analyses. Four time zone predictions are done with SC-ANFIS. For each zone 1000 training data points are taken and 100 data points used for testing and validation. Simulation is carried out for 500 epochs. Here trial and error method it is found that Subtractive clustering method with radius of influence .8 gives best result. For four time zone results are tabulated in Table 4.2. And Figure 4.22 to Figure 4.26 gives detailed plot MFs before training and after training, error curves and prediction accuracy respectively. It can be seen that prediction accuracy very high as compared to neural network method. If the rules criteria is allowed have maximum than ANFIS has infinite prediction ability as mentioned in reference [**20**].

From tabulation it is seen that ANFIS has better performance in terms prediction accuracy and RMSE error. Subtractive clustering based ANFIS generates optimum rules based on SC algorithm. As compared to NNs method proposed earlier ANFIS method performance is better in all respect. Number rules used for trained ANFIS are 6. And prediction accuracy for all case above 90 percentage. And this method provides faster training than NNs methods. Since number of tunable parameters compared to neural network are less, so it would make easy to be implemented in Hardware.

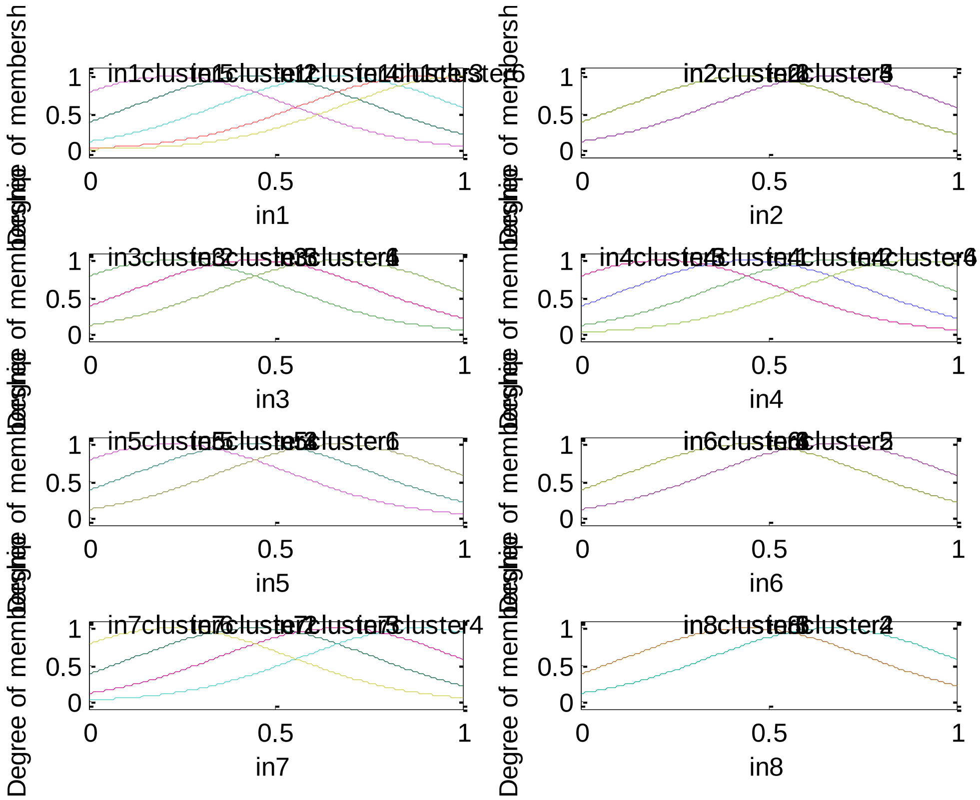

Figure 4. 22 Membership functions before training

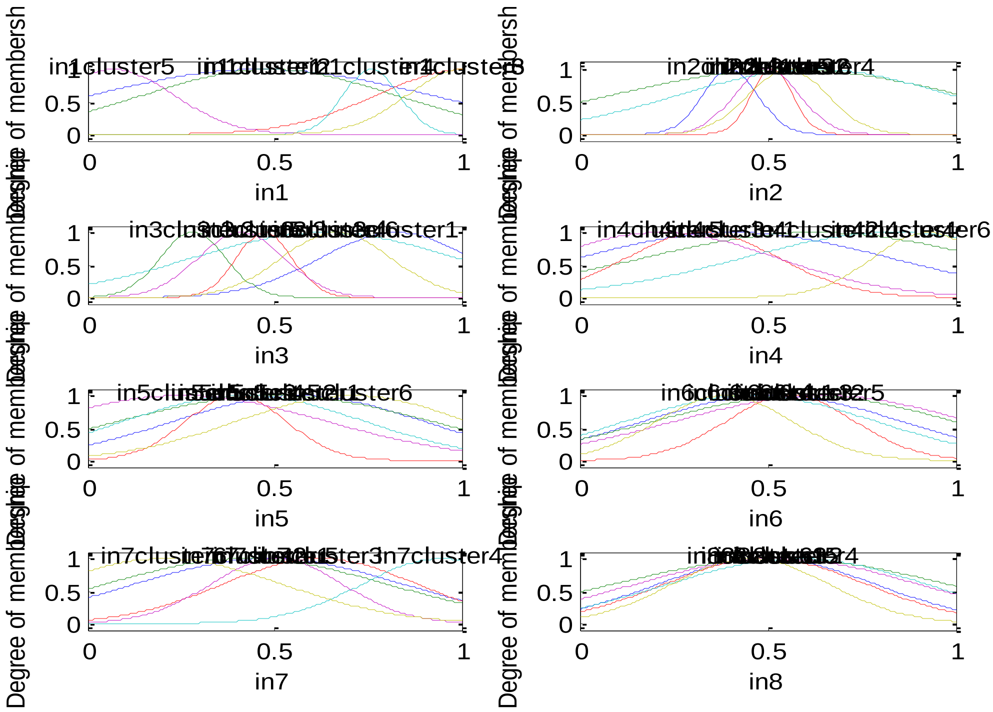

Figure 4. 23 Membership functions after training.

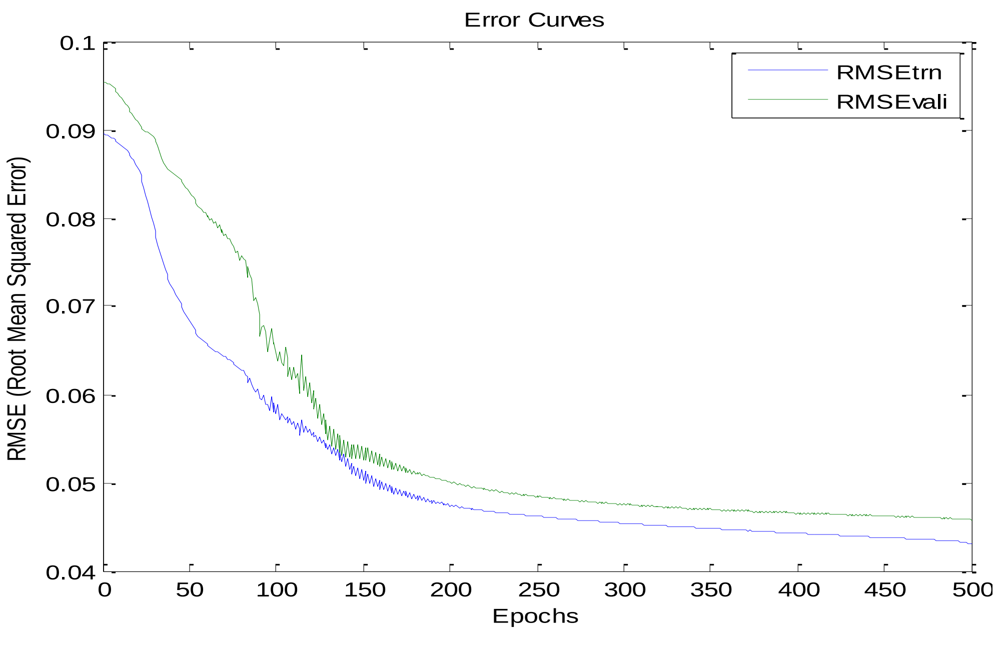

Figure 4. 24 RMSE curves after training and validation.

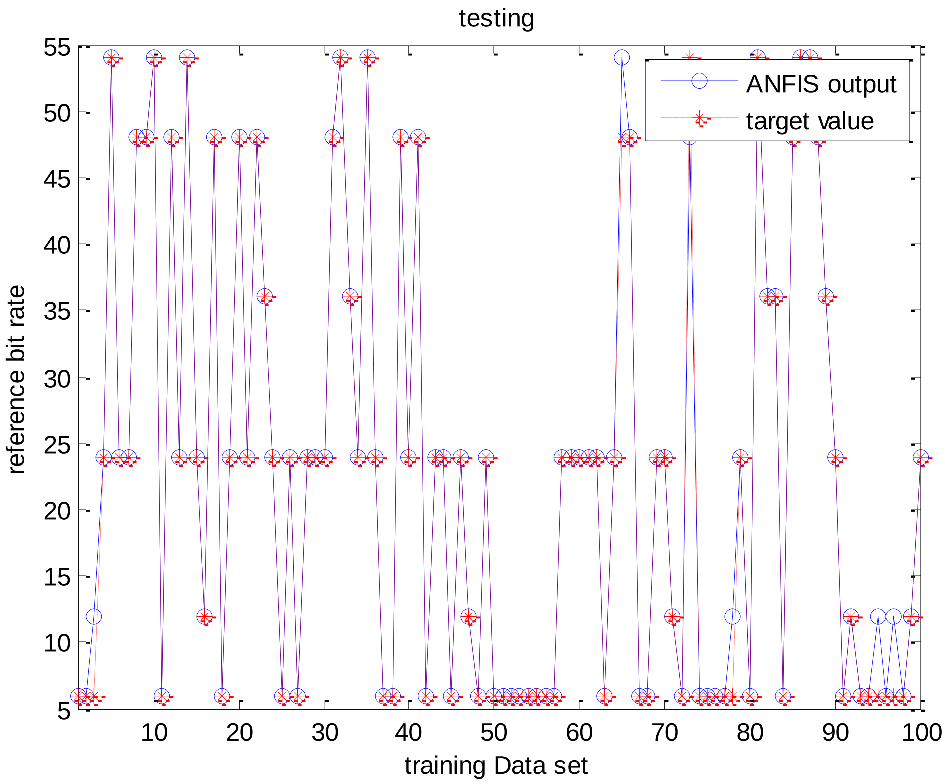

Figure 4. 25 Prediction accuracy of subtractive clustering based ANFIS in training sequence-extended scheme scheme..

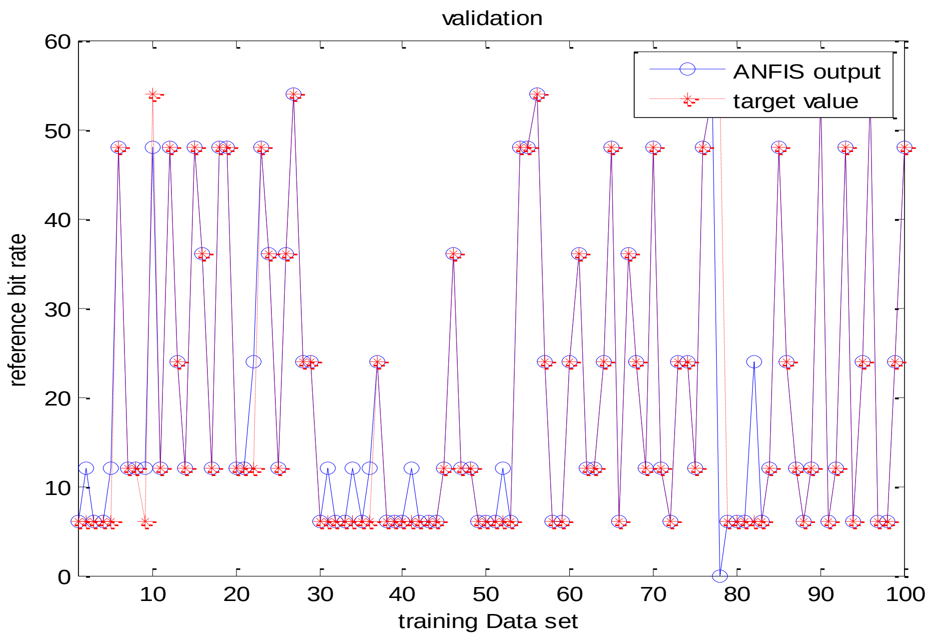

Figure 4. 26 Prediction accuracy of subtractive clustering based ANFIS in validation sequence-extended scheme.

Table 4.2 Performance index of ANFIS and NNs techniques -extended case

Technique	Hidden nodes	Hidden layers	Rules	$RMSE_{trn}$	$RMSE_{val}$	$\lvert \llbracket RMSE - RMSE \rrbracket$	Total number tunable parameters	Prediction accuracy Training	Prediction accuracy Validation
ANFIS-SC		5	12	.0236	.1230	0.0994	158	94	89
ANFIS-SC		5	6	.0328	.1180	0.0852	150	97	91
ANFIS-SC		5	6	0.0573	.0693	.0120	175	94	93
ANFIS-SC		5	32	.0118	.2007	.1889	850	100	88
FF	15	1		.1115	.1646	.0531	158	84	77
Elman	15	1		.1225	.1011	.0124	383	76	73
FTDNN	10	1		.0637	.5947	.531	158	85	78
FF	10-10	2		.0903	.1516	.0613	218	92	88
FF	8-10-10	3		.0642	.1462	.082	300	83	82

4.5 Conclusion

ANFIS has been successfully implemented for data rate prediction to assist cognitive radio both for basic and extended scheme. Here NNs based reference model which was previously implemented in [**4**] is compared. ANFIS based method outperformed NN based technique in all metrics as previously disused. Here dynamic radio configuration is considered as single objective problem and is solved. But practical situation involves multi objective requirement, where it need to predict types of modulation ,frame rate environmental conditions ,etc.. . It is expected that same ANFIS can be used to solve multi objective problem. Cognitive radio devices need to efficiently perceive the necessity for alternating their radio configuration, to evaluate the capabilities of each of the candidate, available configurations and thereupon, to dynamically select the one, in which they must operate. Thesis proposes solution to assist the cognitive radios in the derivation and enforcement of decisions regarding the selection of the desired radio configuration, the one that optimizes its QoS. The proposed solution combines both advantage of neural network and Fuzzy theory. So they have ability create to rules like fuzzy and adaptive learning like NNs. It is seen proposed ANFIS technique is one better candidate to implement cognitive tasks in CR, which can be tried any phase of cognition cycle.

CHAPTER 5

CONCLUSION

5.1 Introduction.

Thesis proposes the ANFIS based data rate prediction for CR in learning. Thesis explores various types of previously used NNs method. Data rate prediction is divided into two schemes that is basic and extended. In basic scheme complexity of problem less and in extended scheme complexity is increased by adding time zone parameter. For the both cases ANFIS techniques were successfully used to predict the data rate.

Chapter discusses the contribution of the thesis, Limitations and scope for future work.

5.2 Contribution of the Thesis

The main purpose of the thesis was to assist data rate prediction for cognitive radio using ANFIS based learning techniques. This thesis uses previous work of [**4**] as reference to implement ANFIS based data rate prediction. Here indirectly capability of radio configuration is estimated. As discussed in chapter 2 the future of wireless communications will be characterized by highly varying environments with multiple available radio access technologies exhibiting diverse features. So in such an unfamiliar landscape, cognitive radio systems are expected to play an exceptional role by adding an inherent ability to perceive, think, decide, learn and adapt to the changing environmental conditions. CR needs learning techniques to act as intelligent radio. As discussed previously there many artificial intelligence techniques to solve this problem. This thesis explores NNs method techniques which are implemented in channel estimating stage of cognition cycle in reference [**4**]. Same procedure is used to put in ANFIS learning technique in channel estimation stage of cognitive radio to predict Data rate of particular radio configuration. By predicting data rate of particular radio configuration proposed ANFIS based technique may facilitate the cognitive terminal in making its decision regarding the configuration in which it should operate, selecting the best among a set of candidate ones.

The ANFIS based technique was successfully implemented to predict data rate in case of basic scheme and extended scheme. Same was compared with neural network based technique which was previously reported. In basic scheme three types ANFIS were used for learning. All three provided better performance in all performance metric with respect to neural network. Conventional ANFIS worked better in accuracy and RMSE error compared all types' neural network method. But it generated huge rule when number inputs were increased and which could not be handled by simulation environment. So to overcome this disadvantage FCM based and subtractive clustering based ANFIS was used.

In extended scheme only subtractive clustering based ANFIS used to predict data rate. This technique provides superior performance compared to previous neural network method in terms of RMSE error and prediction accuracy. The numerical complexity of ANFIS is less than NN. Because SC methods which generates optimum rules reduces mathematical complexity, where as in neural networks number of nonlinear functions and updating weights are very high compared tunable parameters in ANFIS technique.

5.3 Limitations

This thesis is prepared as an extraction of one year research work, which is part of Master of Technology curriculum.

- Training and testing of all the models were conducted offline. Since CR is intelligent device so method has to be devised for online process.
- CR complex intelligent radio, so QoS of CR cannot be optimized by only predicting Data rate, other parameters like modulation type, frame rate and environment conditions must be considered
- Proposed method practical applicability has to be verified.

5.3 Scope for Future work

Proposed ANFIS based technique was successful in only prediction of data rate capability of a specific radio configuration. Capability radio configuration not only depends on data rate, it may include different access technology, modulation type, frame rate etc. So ANFIS based technique must be tuned predict all these capability of radio configuration. So for this different types of hybrid ANFIS must be explored. In extended case only time zone parameter is included but practical situation environmental conditions also affect data rate and other radio capabilities. Problem must be formulised to include other parameters which affect data rate. The prediction was based on assumed scenario but to validate and check the robustness of ANFIS more realistic time series must be considered for training. AS previously said CR sits on SDR so ANFIS based methods feasibility in hardware implantations must be checked.

Bibliography

[1] S Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J,Select. Areas Commun*, vol. 23, pp. 201-220, 2005.

[2] Vijay K.Bhargava, Ekram Hossain, *Cognitive Wireless Communication Networks*, 1st ed., Springer, Ed., 2007.

[3] J. Mitola, "Cognitive Radio an Integrated Agent Architecture for Software Defined," KTH Royal Institute of Technology, Stockholm,Sweden, PhD thesis 2000.

[4] A. Katidiotis, P. Demestichas K. Tsagkaris, "Neural network-based learning schemes for cognitive radio systems," *Computer Communications*, vol. 31, no. 14, pp. 3394-3404, September 2008.

[5] Roy Rubenstein, "Radios Get Smart," *IEEE Spectrum, consumer electronics*, FEBRUARY 2007.

[6] Mubbashar Altaf Khan, Sohib Ahamad, "Decision Making Techniques For COGNITIVE RADIOS," Blekinge Institute of Technology, Master of Science thesis MARCH 2008.

[7] J Mitola, "Cognitive radio: making software radios more personal," *IEEE Pers. Commun*, vol. 6, no. 4, pp. 13-18, 1999.

[8] Bruce A. Fette, *Cognitive Radio Technology*, 1st ed.: Newnes, 2006.

[9] "FCC;ET Docket No 03-222 Notice of proposed rule making and order," December 2003.

[10] James O'Daniell Neel, "Analysis and Design of Cognitive Radio Networks and Distributed Radio Resource Management Algorithms," Virginia Polytechnic Institute and State University, Doctoral Dissertation September 6, 2006.

[11] Mohamed Gafar Ahmed Elnourani, "Cognitive Radio And Game Theory: Overview And Simulation," Blekinge Institute of Technology, Master of Science thesis 2008.

[12] Carl R. Stevenson,Gerald Chouinard,Zhongding Lei,Wendong Hu,Stephen J. Shellhammer,Winston Caldwell, "IEEE 802.22: The First Cognitive Radio Wireless Regional Area Network Standard," *IEEE Standards In Communications*, pp. 130-138, January 2009.

[13] SDR forum. [Online]. http://www.sdrforum.org

[14] D. Raychaudhuri. WINLAB. [Online]. www.winlab.rutgers.edu

[15] The Berkeley Wireless Research Center. [Online]. http://bwrc.eecs.berkeley.edu

[16] Simon Haykin, *Neural Networks,A Comprehensive Foundation*, 2nd ed. South Asia:

Pearson Education and Dorling Kindersley Pvt.Ltd, 1999.

[17] Yu Hen Hu Jenq-Neng Hwang, *Handbook of Neural Network Signal Processing*, 1st ed.: CRC, 2001.

[18] Martin T Hagan,Howard B.Demuth,Mark Beale, *Neural Network Design*, 1st ed.: PWS publishing, 2002.

[19] Serhat -Seker, Emine Ayaz, Erdinc T.urkcan, "Elman’s recurrent neural network applications to conditionmonitoring in nuclear power plant and rotating machinery," *Engineering Applications of Artificial Intelligence*, no. 16, pp. 647-656, 2003.

[20] J.-S. R Jang, "ANFIS: Adaptive-Network-based Fuzzy Inference Systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665-685, May 1993.

[21] Jyh-Shing Roger Jang ,Chuen-Tsai Sun,Eiji Mizutani, *Neuro-Fuzzy and Soft Computing A Computational Approach to Learning and Machine Intelligence*, 1st ed. Delhi: Pearson Education, 2004.

[22] S.M.Fahimifard, M.Homayounifar,M.Sabouhi and A.R.Moghaddamnia, "Comparison of ANFIS,ANN,GARCH and ARIMA Techniques to Exchange Rate Forecasting," *Journal of Applied Sciences*, vol. 9, no. 20, pp. 3641-3651, 2009.

[23] Lim Eng Aik & Yogan S/O Jayakumar, "A Study of Neuro-fuzzy System in Approximation-based Problems," *MATEMATIKA*, vol. 24, no. 2, pp. 113-130, 2008.

[24] B. Fritzke, "Incremental neuro-fuzzy systems," in *Applications of Soft Computing*, San Diego, 27 July - 1 August 1997.

[25] Kuik Sok Ping, Naomie bt Salim, "Optimized Subtractive Clustering for Cluster-Based Compound Selection," in *Proceedings of the 1st International Conference on Natural Resources Engineering & Technology*, Putrajaya, Malaysia, 24-25th July 2006;, pp. 492-499.

[26] Simon Haykin, *Communication Systems*, 3rd ed. Sinapore: Johan Wiley & sons , 1993.

[27] Junhong Nie & Derek Linkens, *Fuzzy-Neural Control*. New Delhi: Printice-Hall India, 1998.

[28] James A Anderson, *An Introduction to Neural Networks*.: Prentice Hall of India, 2002.

[29] H. Demuth, M. Beale, M. Hagan, *Matlab Neural Network Toolbox User’s Guide,Version 5.1.*: The MathWorks Inc., 2007.

Dissemination of research work

IEEE Indexed International Conference Publication.

1. Shrishail Hiremath and Prof. S.K.Patra "Transmission Rate Prediction for Cognitive Radio Using Adaptive Neural Fuzzy Inference System" accepted at International Conference on Industrial and Information Systems (ICIIS 2010), NITK, Surathkal, Karnataka.